

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : C12N 15/52, 9/00, C12Q 1/25	A2	(11) International Publication Number: <b>WO 00/60091</b> (43) International Publication Date: 12 October 2000 (12.10.00)
(21) International Application Number: PCT/US00/09154 (22) International Filing Date: 6 April 2000 (06.04.00) (30) Priority Data: 60/127,977          6 April 1999 (06.04.99)          US (71) Applicant: OKLAHOMA MEDICAL RESEARCH FOUNDATION [US/US]; 825 N.E. Thirteenth Street, Oklahoma City, OK 73104 (US). (72) Inventors: SAUER, Brian, Lee; 601 Robert S. Kerr Avenue, Apartment 109, Oklahoma City, OK 73102 (US). RUFER, Andreas, Walter; 125 N.W. Fifteenth Street, Apartment 604, Oklahoma City, OK 73103 (US). (74) Agents: PABST, Patrea, L. et al.; Amall Golden & Gregory, LLP, 2800 One Atlantic Center, 1201 West Peachtree Street, Atlanta, GA 30309-3450 (US).	(81) Designated States: AU, CA, IL, JP, ZA, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(54) Title: METHOD FOR SELECTING RECOMBINASE VARIANTS WITH ALTERED SPECIFICITY		
(57) Abstract  <p>Disclosed are variants of Cre recombinase that have broadened specificity for the site of recombination. Specifically, the disclosed variants mediate recombination between sequences other than the loxP sequence and other lox site sequences on which wild type Cre recombinase is active. In general, the disclosed Cre variants mediate efficient recombination between lox sites that wild type Cre can act on (referred to as wild type lox sites), between variant lox sites not efficiently utilized by wild type Cre (referred to as variant lox sites), and between a wild type lox site and a variant lox site. Also disclosed are methods of recombining nucleic acids using the disclosed Cre variants. For example, the disclosed Cre variants can be used in any method or technique where Cre recombinase (or other, similar recombinases such as FLP) can be used. In addition, the disclosed Cre variants allow different alternative recombinations to be performed since the Cre variants allow much more efficient recombination between wild type lox sites and variant lox sites. Control of such alternative recombination can be used to accomplish more sophisticated sequential recombinations to achieve results not possible with wild type Cre recombinase.</p>		

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**Description**

5

10

15

20

25

30

35

40

45

50

55

**METHOD FOR SELECTING RECOMBINASE VARIANTS  
WITH ALTERED SPECIFICITY  
BACKGROUND OF THE INVENTION**

Recombinases, integrases and resolvases (collectively referred to herein as recombinases) mediate the site-specific recombination of DNA. These recombinases were first identified in phage that integrate into host chromosomes. Such integration allows the phage to remain latent in the cell as a prophage.

Site-specific recombinases catalyze conservative DNA rearrangements at specific target sequences. The 38 kDa Cre recombinase (cyclization recombination), derived from the bacteriophage P1, is a well characterized and widely used enzyme of the Integrase family (reviewed by Sauer, *Methods*, 14:381-392 (1998)). Cre plays two essential roles in the life cycle of P1: First, it provides a host-independent mechanism for P1's genome cyclization after infection, which can be important when the recombination system of the host is compromised. Second, Cre resolves dimerized P1 prophage plasmids to guarantee proper segregation during cell division.

Cre acts on a 34 bp sequence located on both ends of the linear P1 genome, that is called *loxP* (locus of crossover of P1; Sternberg and Hamilton, *J. Mol. Biol.*, 150:467-486 (1981)), *loxP* consists of two 13 bp inverted repeats flanking a non-palindromic 8 bp core that defines the assigned direction of the sequence (as shown on the upper part of Figure 1). Depending on this direction, recombination catalyzed by Cre leads to excision of insertion of DNA flanked by *loxP* sites orientated in the same direction (indicated by *loxP*<sup>2</sup>), but leads to inversion when oriented in the opposite direction (Figure 2).

In general, Cre-recombination involves the following four events: (i) DNA binding, (ii) synapsis (as defined below), (iii) cleavage, and (iv) strand exchange. To study this process in greater detail, mutants defective for each step have been isolated using several screening procedures (Wierzbicki et al., *J. Mol. Biol.*, 195:785-794 (1987)). In addition, the crystal structure of Cre complexed with an artificial suicide substrate has been recently resolved, providing additional insights into site-specific recombination (Guo et al.,

5 *Nature*, 389:40-46 (1997)). From these studies, the following has been  
proposed: Four interacting Cre molecules are necessary for recombination  
between two *lox* sites, with each enzyme binding one inverted repeat plus the  
10 two outermost bp of the non-symmetric core region (DNA binding). This leads  
5 to the formation of a clamp, allowing DNA contacts in the major, as well as in  
the minor groove. In the step referred to as synapsis, the two *lox* sites with the  
bound Cre molecules, are aligned in parallel leading to an approximate 100°  
15 bending of the DNA. In the following step of strand cleavage, one of the two  
Cre molecules on each *lox* site causes a staggered cut in the core region, as  
10 indicated by the vertical arrows in Figure 1. This leads to a 6 bp 5' overhang  
and a covalent 3' phosphotyrosine linkage between the catalytic residue tyrosine  
324 of Cre and the guanine (position 4) at the cleaving site of *loxP*. The created  
20 phosphotyrosine intermediate is thought to provide the energy for the reaction,  
thereby explaining why Cre does not require an external energy source. In the  
25 next step, the first strand is exchanged between the two nicked *lox* sites,  
creating an intermediate, named Holliday structure (Sigal and Alberts, *J. Mol.*  
*Biol.*, 71:789-793 (1972)). Of note, this first strand exchange is asymmetric,  
since the bottom strand (Figure 1) is always exchanged first (Hoess et al., *Proc.*  
30 *Natl. Acad. Sci. USA*, 84:6840-6844 (1987)). During the final step, the second  
20 strand is exchanged and Cre released from its substrate.

Because of the simplicity and the ability of Cre to function in yeast and  
35 mammalian cells (Sauer, B., *Mol. Cell. Biol.*, 7:2087-2096 (1987); Sauer and  
Henderson, *Proc. Natl. Acad. Sci. USA*, 85:5166-5170 (1988), Sauer and  
Henderson, *Nucl. Acids Res.*, 17:147-161 (1989), and Sauer and Henderson, *The*  
25 *New Biologist*, 2:441-449 (1990), Cre assisted site-specific recombination has  
become an important tool for efficient, specific, and conditional manipulations  
40 of eukaryotic genomes (Lakso et al., *Proc. Natl. Acad. Sci. USA*, 89:6232-6236  
(1992)); Kilby et al., *Genet.*, 9:413-421 (1993); Sauer, B., *Meth. enzymol.*,  
225:890-900 (1993); Kühn et al., *Science*, 269:1427-1429 (1995); Metzger et  
45 al., *Pro. Natl. Acad. Sci. USA*, 92:6991-6995 (1995).

However, there are some inconveniences for the successful use of Cre-  
50 related technologies, that include the following: (i) *lox* sites need to be

introduced by homologous recombination at the desired region into the genome before Cre can be used, (ii) the frequency of correct site-specific recombination due to Cre expression is not 100%, and consequently, (iii) selectable markers are necessary in most strategies involving Cre for genome manipulation in higher eukaryotes. These markers, e.g. *neo* or TK, may introduce problems in subsequent studies, particular in those related to animal development. The number of available selectable markers that can be used is limited also. Additional site-specific recombinases that also function efficiently in eukaryotic systems, but recognize different sites from *lox* would be helpful. Similar inconveniences limit the usefulness of other recombinases.

Therefore, it is an object of the present invention to provide a method of identifying variant recombinases that can mediate recombination between variant recombination sites.

It is another object of the present invention to provide variant recombinases that can mediate recombination between variant recombination sites.

It is another object of the present invention to provide a method of recombining nucleic acid molecules *in vitro* and *in vivo*.

It is another object of the present invention to provide Cre variants that recognize variant recombination sites.

#### BRIEF SUMMARY OF THE INVENTION

Disclosed is a method for identifying variant forms of recombinases that can mediate recombination between variant recombination sites. The method involves producing mutant recombinases and testing the mutant recombinases with specially designed constructs. The constructs contain variant recombination sites that are not recognized by non-mutant recombinase but will undergo recombination in the presence of a mutant recombinase with altered specificity. Recombination at the variant recombination sites can be monitored or detected by any suitable means. It is preferred that recombination is detected by screening or selection based on the expression or lack of expression of a reporter gene. This can be accomplished by using constructs containing a reporter gene associated with the variant recombination sites such that the

5 reporter gene is rearranged or deleted, or a spacer sequence interrupting the  
reporter gene is rearranged or deleted, as a result of recombination at the  
10 recombination sites. Recombination of such constructs will result in a loss of  
expression of the reporter gene, where the construct contained a functional  
5 reporter gene, or in a gain in expression of the reporter gene, where the  
construct contained a non-functional reporter gene.

15 The disclosed method also involves determining whether a variant  
recombinase retains the ability to mediate recombination at recombination sites  
recognized by non-variant recombinase. This can be accomplished by using  
10 constructs containing recombination sites recognized by non-variant  
recombinase. Recombination at these recombination sites can be monitored or  
20 detected by any suitable means. It is preferred that recombination is detected by  
screening or selection based on the expression or lack of expression of a  
reporter gene. This can be accomplished by using constructs containing a  
25 reporter gene associated with the recombination sites recognized by non-variant  
recombinase such that the reporter gene is rearranged or deleted, or a spacer  
sequence interrupting the reporter gene is rearranged or deleted, as a result of  
30 recombination at the recombination sites. Recombination of such constructs  
will result in a loss of expression of the reporter gene, where the construct  
20 contained a functional reporter gene, or in a gain in expression of the reporter  
gene, where the construct contained a non-functional reporter gene.

35 When variant recombinases are tested for activity on both variant  
recombination sites and recombination sites recognized by non-variant  
recombinase in the same system or at the same time, it is preferred that two  
25 different reporter genes, which can be separately detected or monitored, be  
used. In this case, the first reporter gene can be associated with the variant  
40 recombination sites and the second reporter gene can be associated with  
recombination sites recognized by non-variant recombinase.

45 Recombination between two recombination sites requires (1) that the  
30 recombinase recognize the sites as recombination sites, and (2) that the  
sequences of the two sites is sufficiently similar. It has been discovered that  
recombination between two recombination sites (both of which are recognized

5 by a recombinase) can be substantially reduced or prevented by using different  
compatibility sequences for the recombination sites (the recognition sequences  
can also differ if the recombinase can recognize different sequences). Thus, it is  
10 also preferred that the variant recombination sites be made incompatible with  
5 the recombination sites recognized by non-variant recombinase by using  
different compatibility sequences for the two sets of recombination sites.

Compatibility sequences in a recombination site are those sequences in  
15 the recombination site, other than the sequences required for recognition of the  
site by the recombinase, that must be similar in a pair of recombination sites for  
10 recombination to occur between them. Many recombination sites contain  
repeats of a characteristic sequence separated by spacer sequences. In such  
20 recombination sites, the spacer sequences are generally compatibility sequences  
and the repeats (or parts of the repeats) are recognition sequences.

Recombinases require specific recognition sequences but allow wide variation  
25 in compatibility sequences. Thus, recombination sites that are recognized by a  
given recombinase but are incompatible with each other can be readily designed  
using the disclosed principles.

Also disclosed are variant recombinases made or identified by the  
30 disclosed method that have broadened specificity for the site of recombination.  
20 Specifically, the disclosed variants mediate recombination between sequences  
other than recombination sites on which the wild type recombinase is active. In  
35 general, the disclosed recombinase variants can mediate efficient recombination  
between recombination sites that wild type recombinase can act on (referred to  
as wild type recombination sites), between variant recombination sites not  
25 efficiently utilized by wild type recombinase (referred to as variant  
40 recombination sites), and between a wild type recombination site and a variant  
recombination site.

Also disclosed are methods of recombining nucleic acids using the  
45 disclosed variant recombinases. For example, the disclosed variant  
30 recombinases can be used in any method or technique where wild type  
recombinases can be used. In addition, the disclosed variant recombinases  
allow different alternative recombinations to be performed since the variant

5 recombinases can allow much more efficient recombination between wild type  
recombination sites and variant recombination sites. Control of such alternative  
recombination can be used to accomplish more sophisticated sequential  
10 recombinations to achieve results not possible with wild type recombinases.

5 The disclosed variant recombinases also allow recombination at specific  
genomic sites without the need to first introduce a recombination site.

Also disclosed are variants of Cre recombinase that have broadened  
15 specificity for the site of recombination. Specifically, the disclosed variants  
mediate recombination between sequences other than the loxP sequence and  
10 other lox site sequences on which wild type Cre recombinase is active. In  
general, the disclosed Cre variants mediate efficient recombination between lox  
20 sites that wild type Cre can act on (referred to as wild type lox sites), between  
variant lox sites not efficiently utilized by wild type Cre (referred to as variant  
lox sites), and between a wild type lox site and a variant lox site. Also disclosed  
25 are methods of recombining nucleic acids using the disclosed Cre variants. For  
example, the disclosed Cre variants can be used in any method or technique  
where Cre recombinase (or other, similar recombinases such as FLP) can be  
used. In addition, the disclosed Cre variants allow different alternative  
30 recombinations to be performed since the Cre variants allow much more  
20 efficient recombination between wild type lox sites and variant lox sites.  
Control of such alternative recombination can be used to accomplish more  
sophisticated sequential recombinations to achieve results not possible with  
35 wild type Cre recombinase.

#### BRIEF DESCRIPTION OF THE DRAWINGS

25 Figure 1 is a comparison of three different lox sites. loxP is the original  
40 recombination site for Cre recombinase. loxK1 and loxK2 are variant lox sites.

Figure 2 is a diagram of two different forms of construct and the  
resulting recombination products.

45 Figure 3 is a diagram of an example of a random mutagenesis using  
30 DNA shuffling.

Figure 4 is a diagram of the selection plasmid for loxK2 recombination,  
pBS584. Recombination of two loxK2 sites by a potent Cre mutant will result

5 in the excision of *EGFP* and the transcriptional terminator *rrnBT<sub>1</sub>T<sub>2</sub>*.  
Subsequently, *neo* transcription can take place, rendering *E. coli* resistant to  
kanamycin. Note that the promoter (pRSV) even though of eukaryotic origin  
10 was shown to be functional in *E. coli* (Antonucci et al., *J. Biol. Chem.*,  
5 264:17656-17659 (1989)).

Figure 5 depicts gels of nucleic acid fragments and PCR products  
generated during the DNA shuffling process.

15 Figure 6 is a diagram of plasmid pBAD33 used for expression of mutant  
*cre* pools.

10 Figure 7 is a diagram of the construction of selection plasmids pBS568  
and pBS569.

20 Figure 8 is a diagram of the construction of selection plasmids pBS583  
and pBS584.

Figure 9 is a diagram of control plasmid pBS613.

15 Figure 10 is a diagram of the construction of screening plasmids pBS601  
25 and pBS602.

Figure 11 is a diagram of examples of basic types of constructs useful in  
the disclosed method. These types of constructs are: (1) interrupted constructs  
30 where the gene is interrupted by a nucleic acid segment (which is flanked by  
20 recombination sites) that is deleted during recombination, (2) flanked constructs  
where the gene as a unit is flanked by recombination sites and the gene is  
35 deleted by recombination, and (3) inverted constructs where a portion of the  
gene is on an inverted nucleic acid segment and recombination causes the  
segment to invert and reconstitute the intact gene. The type of recombination is  
25 indicated in parentheses.

40 Figures 12A, 12B, and 12C are diagrams of examples of constructs and  
their expected recombination when used in the disclosed method. Figure 12A  
shows examples of deletion constructs (flanked and interrupted). Figure 12B  
shows examples of inverted constructs. Figure 12C shows examples of  
45 30 constructs that combine through recombination to reconstitute an intact gene.

Figure 13 is a diagram showing the identified amino-acid changes in the  
50 six selected Cre mutants are listed according to their position in the protein's

secondary structure (silent mutations in parenthesis). Only one amino acid change, E262G, is common to all mutants with remarkably increased *loxK2* activity (R3M1, 2, 3, 5, and 6), suggesting that this mutation is essential for the observed phenotype.

Figure 14A is a table comparing recombination frequencies *in vivo* obtained with a variety of *lox* sites altered at positions 11 and 12. Figure 14B is a table comparing recombination frequencies *in vivo* obtained with identical and mixed *lox* sites. Wild type Cre and five different mutant enzymes were tested for their performance on different *lox*<sup>2</sup> substrates, as indicated. Given are the obtained percentages of recombination *in vivo* based on the described negative selection.

Figure 15 is a table comparing recombination frequencies *in vitro* obtained with a variety of *lox* sites altered at positions 11 and 12.

Figure 16 is a graph of percent of various Cre recombinases (wt, G, GA, GN, GS, R3M3) bound to various *lox* sites (*loxP*, *loxK2*, *loxK1*).

Figure 17 shows wildtype and target FRT sites.

Figure 18 shows the strategy for selection of altered specificity FLP mutants.

Figure 19 shows an alternate target mutant FRT site. The design and rationale for design of the target mutant FRT site is as described in Figure 17, but the mutant FRT-M2 site differs from FRT-M by carrying a different mutational alteration in the repeat elements.

#### DETAILED DESCRIPTION OF THE INVENTION

Disclosed is a method for identifying variant forms of recombinases that can mediate recombination between variant recombination sites. The method involves producing mutant recombinases and testing the mutant recombinases with specially designed constructs. The constructs contain variant recombination sites that are not recognized by non-mutant recombinase but will undergo recombination in the presence of a mutant recombinase with altered specificity. The disclosed method also involves determining whether a variant recombinase retains the ability to mediate recombination at recombination sites recognized by non-variant recombinase.

When variant recombinases are tested for activity on both variant recombination sites and recombination sites recognized by non-variant recombinase in the same system or at the same time, it is preferred that two different reporter genes which can be separately detected or monitored be used.

In this case, a first reporter gene can be associated with the variant recombination sites and a second reporter gene can be associated with recombination sites recognized by non-variant recombinase. It is also preferred that the variant recombination sites be made incompatible with the recombination sites recognized by non-variant recombinase by using different compatibility sequences for the two sets of recombination sites. This allows separate assessment of the ability of a variant recombinase to mediate recombination between variant recombination sites and recombination sites recognized by non-variant recombinase.

Also disclosed are variant recombinases made or identified by the disclosed method that have broadened specificity for the site of recombination. Also disclosed are methods of recombining nucleic acids using the disclosed variant recombinases. For example, the disclosed variant recombinases can be used in any method or technique where wild type recombinases can be used. In addition, the disclosed variant recombinases allow different alternative recombinations to be performed since the variant recombinases can allow much more efficient recombination between wild type recombination sites and variant recombination sites. Control of such alternative recombination can be used to accomplish more sophisticated sequential recombinations to achieve results not possible with wild type recombinases.

Also disclosed are variants of Cre recombinase that have broadened specificity for the site of recombination. Specifically, the disclosed variants mediate recombination between sequences other than the loxP sequence and other lox site sequences on which wild type Cre recombinase is active. Preferred forms of the disclosed Cre variants have the amino acid sequence SEQ ID NO:1 (top sequence, Table 11) altered by one or more amino acid substitutions, deletions, or insertions, where the glutamic acid at amino acid 262 has been substituted with an amino acid other than glutamic acid, and where the

Cre variant recognizes (that is, mediates recombination at) a variant lox recombination site. Useful Cre variants include proteins that recognize a variant lox recombination site and have the amino acid sequence SEQ ID NO:1 altered by substitution of the glutamic acid at amino acid 262 with an amino acid other than glutamic acid and one or more of the following amino acid substitutions: isoleucine at amino acid 16, alanine at amino acid 29, glutamine at amino acid 101, glycine at amino acid 138, asparagine at amino acid 189, serine at amino acid 198, glutamine at amino acid 220, glutamine at amino acid 223, isoleucine at amino acid 227, glycine at amino acid 254, arginine at amino acid 255, glutamine at amino acid 284, leucine at amino acid 307, and serine at amino acid 316. Preferred amino acid substitutions at amino acid position 262 include alanine, tryptophan, or glycine.

Examples of preferred Cre variants include proteins having the amino acid sequence SEQ ID NO:1 altered by substitutions E262G and D189N; proteins having the amino acid sequence SEQ ID NO:1 altered by substitutions E262G and T316S; proteins having the amino acid sequence SEQ ID NO:1 altered by substitutions E262G and D29A; proteins having the amino acid sequence SEQ ID NO:1 altered by substitutions E262G, V16I, D189N, G198S, R223Q, Q255R, and P307L; proteins having the amino acid sequence SEQ ID NO:1 altered by substitution E262G; proteins having the amino acid sequence SEQ ID NO:1 altered by substitution E262A; and proteins having the amino acid sequence SEQ ID NO:1 altered by substitution E262W. The substitutions above are listed using the convention where the first letter is the original amino acid (in single letter amino acid code), the number is the amino acid position in the protein (in this case, using the positions of wild type Cre (SEQ ID NO:1)), and the last letter is the new amino acid (in single letter amino acid code). All of these Cre variants recognize both wild type lox sites and variant lox sites with an inverted repeat sequence NNNACNNCGTATA (SEQ ID NO:2).

The disclosed Cre variants recognize variant lox recombination sites. Preferred variant lox sites are variant lox sites recognized by the Cre variant but not recognized by wild type Cre. Examples of useful variant lox sites include sites having two 13 base pair inverted repeats flanking 8 base pairs, where one

of the inverted repeats has the sequence NNNACNNCGTATA (SEQ ID NO:2); sites having the sequence N<sub>1</sub>N<sub>2</sub>N<sub>3</sub>ACN<sub>4</sub>N<sub>5</sub>CGTATANNNNNNNNTATA CGN<sub>5</sub>'N<sub>4</sub>'GTN<sub>3</sub>'N<sub>2</sub>'N<sub>1</sub>' (SEQ ID NO:3), where N<sub>1</sub>', N<sub>2</sub>', N<sub>3</sub>', N<sub>4</sub>', and N<sub>5</sub>' are complementary to N<sub>1</sub>, N<sub>2</sub>, N<sub>3</sub>, N<sub>4</sub>, and N<sub>5</sub>, respectively; sites having the sequence N<sub>1</sub>N<sub>2</sub>N<sub>3</sub>ACN<sub>4</sub>N<sub>5</sub>CGTATANNNNNNNNTATACGN<sub>5</sub>'N<sub>4</sub>'GN<sub>3</sub>'N<sub>2</sub>'N<sub>1</sub>' (SEQ ID NO:3), where N<sub>4</sub>N<sub>5</sub> are AA, TC, GT, TG, GG, or CC; and sites having the sequence GATACAACGTATATACCTTCTATACGTTGTAT (SEQ ID NO:4).

Also disclosed is a method for producing site-specific recombination of DNA in cells using the disclosed Cre variants. DNA sequences comprising first and second lox sites are introduced into cells and contacted with a Cre variant, thereby producing recombination at the lox sites. As with wild type Cre, the location and orientation of the lox sites determines the nature of the recombination.

As used herein, the expression "site-specific recombination" refers to three different types of recombination events:

1. deletion of a pre-selected DNA segment flanked by recombination sites,
2. inversion of the nucleotide sequence of a pre-selected DNA segment flanked by recombination sites, and
3. reciprocal exchange of DNA segments proximate to recombination sites located on different DNA molecules.

It is to be understood that this reciprocal exchange of DNA segments can result in an integration event if one or both of the DNA molecules are circular. "Nucleic acid segment" refers to a linear segment of single- or double-stranded nucleic acid, which can be derived from any source. The segment may be a fragment consisting of the segment or a segment within a larger nucleic acid fragment or molecule. The expression "nucleic acid in eukaryotic cells" includes all nucleic acid present in eukaryotic cells. The expression "nucleic acid in yeast" includes all nucleic acid present in yeast cells. "DNA segment" refers to a linear segment of single- or double-stranded deoxyribonucleic acid (DNA), which can be derived from any source. The expression "DNA in

5 eukaryotic cells" includes all DNA present in eukaryotic cells. The expression  
"DNA in yeast" includes all DNA present in yeast cells. As used herein, a  
"gene" is intended to mean a DNA segment which is normally regarded as a  
10 gene by those skilled in the art. The expression "regulatory molecule" refers to  
5 a polymer of ribonucleic acid (RNA) or a polypeptide which is capable of  
enhancing or inhibiting expression of a gene.

15 "Regulatory nucleotide sequence," as used herein, refers to a nucleotide  
sequence located proximate to a gene whose transcription is controlled by the  
regulatory nucleotide sequence in conjunction with the gene expression  
10 apparatus of the cell. Generally, the regulatory nucleotide sequence is located 5'  
to the gene. The expression "nucleotide sequence" refers to a polymer of DNA  
or RNA, which can be single- or double-stranded, optionally containing  
20 synthetic, non-natural, or altered nucleotides capable of incorporation into DNA  
or RNA polymers. As used herein, a "regulatory nucleotide sequence" can  
15 include a promoter region, as that term is conventionally employed by those  
skilled in the art. A promoter region can include an association region  
recognized by an RNA polymerase, one or more regions which control the  
effectiveness of transcription initiation in response to physiological conditions,  
30 and a transcription initiation sequence. "Gene product" refers to a polypeptide  
20 resulting from transcription, translation, and, optionally, post-translational  
processing of a selected DNA segment.

#### Materials

##### A. Recombinases

35 Recombinases suitable for use in the disclosed method include any  
25 enzyme that mediates recombination at specific sites. This includes enzymes  
identified as recombinases as well as other enzymes that function to produce  
40 recombination such as integrases and resolvases. As used herein, recombination  
at specific sites does not refer only to recombination at completely defined  
sequences. Rather, a recombinase is considered to mediate recombination at  
45 specific sites when the sites of recombination are limited in some way by  
30 sequence. For example, wild type Cre recombinase mediates recombination  
between sites having the sequence  $N_1N_2N_3$ ACTTCGTATANNNNNNNNT

5 ATACGAAGTN<sub>3</sub>'N<sub>2</sub>'N<sub>1</sub>', which includes both specific and non-specific sequences. The sequences ACTTCGTATA and TATACGAAGT (an inverted repeat of the first sequence) are recognized by the Cre recombinase. The non-specific sequences (positions with "N" in the recognition sequence), although  
10 5 not limited in sequence, must be compatible with the non-specific sequences of the partner recombination site in order for recombination to be efficient. The recombination sites need not have any particular number of specific nucleotides. All that is required is some constraint on the sequence of the site such that the  
15 recombination is limited to recombination at some set of sites.

10 Examples of recombinases that can be used in the disclosed method include Cre recombinase, FLP recombinase, Beta recombinase of pSM19035 (Diaz et al., *J Biol Chem* 274: 6634-6640 (1999)), Int recombinases (Nunes-Döby et al., *Nucleic Acids Res.* 26:391-406 (1998)), and resolvases (Hallet et al., *FEMS Microbiol Rev.* 21: 157-178 (1997); Oram et al., *Curr Biol.* 5: 1106-  
15 1109 (1995); Mondragon, *Structure* 3: 755-758 (1995)).

#### 25 B. Recombination Sites

Recombination sites are locations within a nucleic acid where recombination mediated by a recombinase takes place. Recombination sites generally include specific sequences, referred to as recognition sequences,  
30 20 through which the recombinase recognizes a given nucleotide sequence as a recombination site. Different recombinases generally recognize different recognition sequences. Recombination between two recombination sites requires (1) that the recombinase recognize the sites as recombination sites, and (2) that the sequences of the two sites are sufficiently similar. It has been  
35 25 discovered that recombination between two recombination sites (both of which are recognized by a recombinase) can be substantially reduced or prevented by using different compatibility sequences for the recombination sites (the recognition sequences can also differ if the recombinase can recognize different sequences). Thus, it is also preferred that the variant recombination sites be  
40 30 made incompatible with the recombination sites recognized by non-variant recombinase by using different compatibility sequences for the two sets of recombination sites. Compatibility sequences in a recombination site are those

5 sequences in the recombination site, other than the sequences required for  
recognition of the site by the recombinase, that must be similar in a pair of  
recombination sites for recombination to occur between them. Generally,  
10 recombinases require specific recognition sequences but allow wide variation in  
5 compatibility sequences. Thus, recombination sites that are recognized by a  
given recombinase but are incompatible with each other can be readily designed  
using the disclosed principles.

15 It should be understood that, for a given recombinase site or for a given  
recombinase, whether a given base position in the recombination site is a  
20 recognition sequence base or a compatibility sequence base may depend on  
other sequences in the recombination site. For example, a particular base may  
function as a compatibility sequence base in a recombination site having one  
sequence while the same base may function as a recognition sequence base in a  
25 recombination site having a different sequence. It should also be understood  
that recognition sequences and compatibility sequences do not necessarily occur  
in blocks within a recombination site. That is, recognition sequence base and  
compatibility sequence bases may be interspersed in a given recombination site.  
30 As discussed below, what is and is not a recognition sequence or a compatibility  
sequence in a given recombination site is determined functionally.

20 The disclosed variant recombination sites and the variant recombinases  
that can act on them allow more freedom in the selection of sites of  
recombination. In particular, the disclosed variant recombinases can allow  
35 amino acid changes in a protein of interest while retaining the ability to  
recombine at a given site.

#### 25 1. Recognition Sequences

40 Recognition sequences are regions within a recombination site that must  
have a specific sequence, or defined range of sequences, for the cognate  
recombinase to recognize the recombination site. Recognition sequences in a  
recombination site need not be contiguous. Thus, required nucleotides  
45 dispersed in a recombination site are collectively considered recognition  
30 sequences. Nucleic acid segments can be said to have a defined range of  
sequences when every nucleotide position in the nucleic acid segment(s) is

5 limited to one, two, or three nucleotide bases. That is, so long as a nucleotide  
position cannot have one of the possible nucleotide bases, that position has a  
defined range of sequence. For example, a nucleotide sequence ATRVBYGC  
10 has a defined range of sequences since each nucleotide position has at least one  
5 limitation. Standard nomenclature for nucleic acid sequences is used herein.  
Thus, in this example, R represents A or G; V represents A, C, or G; B  
represents C, G, or T; and Y represents C or T.

15 Recognition sequences for recombinases are known or can be  
determined through routine analysis. In general, recognition sequences can be  
10 determined by varying the sequence of recombination sites and determining if  
recombination between the sites can still occur. For this purpose, the pair of  
20 sites to be recombined should be identical. That is, the same sequence changes  
should be made to both sites. This eliminates any incompatibility effect  
between the recombination sites. If recombination is eliminated or significantly  
15 reduced when a specific nucleotide is changed, then that nucleotide is required  
for recognition of the recombination site by the recombinase.

25 Examples of dissection of the critical sequences in recombination sites  
of recombinases are described by Hoess et al., *Nucleic Acids Res.* 14:2287-2300  
30 (1986) (involving P1 recombinase); Sauer B., *Nucleic Acids Res.*, 24:4608-4613  
20 (1996) (involving Cre recombinase); Lee and Saito, *Gene* 216(1):55-65 (1998)  
(involving Cre recombinase); and Umlauf and Cox, *EMBO J* 7(6):1845-52  
35 (1988) (involving FLP recombinase). Similar techniques can be used to  
determine the recognition sequences of other recombinases.

## 2. Compatibility Sequences

25 Compatibility sequences are regions in a recombination site that must be  
similar in a pair of recombination sites for recombination to occur between  
40 them. In general, the sequence of recombination sites must be similar for  
recombination to occur between them. Examples of compatibility sequences  
are spacer sequences between repeats in recombination sequences. All or some  
45 of the nucleotides in the recognition sequences for a recombination site may be  
30 involved in compatibility. For example, where some degeneracy of the  
recognition sequences is allowed, similar recognition sequences may be

5 required in a pair of recombination sites for recombination to occur between  
them. Thus, compatibility between recombination sites can be affected by using  
different sequences in the compatibility sequences other than the sequences  
10 required for recognition of the site by the recombinase (that is, recognition  
5 sequences), compatibility sequences that are part of the recognition sequences,  
or both. It is preferred that compatibility between recombination sites be altered  
by using different sequences in the compatibility sequences other than the  
15 sequences required for recognition of the site by the recombinase.

Compatibility sequences for many recombinases are known or can be  
10 determined through routine analysis. In general, compatibility sequences can be  
easily determined by varying the sequence of recombination sites and  
20 determining if recombination between the sites can still occur. For this purpose,  
only one of the sites in the pair of sites to be recombined should be altered.  
That is, the same sequence changes should not be made to both sites. This  
15 isolates incompatibility effect between the recombination sites. Further, only  
those nucleotide positions that are not a part of the recognition sequence of the  
25 site should be altered to avoid recognition problems. If recombination is  
eliminated or significantly reduced when a specific nucleotide is changed, then  
30 that nucleotide is required for compatibility of the recombination site.

20 Examples of dissection of the critical sequences in recombination sites  
of recombinases are described by Hoess et al., *Nucleic Acids Res.* 14:2287-2300  
(1986) (involving P1 recombinase); Sauer B., *Nucleic Acids Res.*, 24:4608-4613  
35 (1996) (involving Cre recombinase); Lee and Saito, *Gene* 216(1):55-65 (1998)  
(involving Cre recombinase); and Umlauf and Cox, *EMBO J* 7(6):1845-52  
25 (1988) (involving FLP recombinase). Similar techniques can be used to  
determine the compatibility sequences of other recombinases.

Recognition and compatibility sequences can be further understood  
using Cre recombination sites as an example. Wild type Cre recombinase  
mediates recombination between sites having the sequence  
45 30  $N_1N_2N_3ACTTCGTATANNNNNNNTATACGAAGTN_3'N_2'N_1'$ , which  
includes both specific and non-specific sequences (that is, recognition  
sequences and compatibility sequences, respectively). The sequences

5 ACTTCGTATA and TATACGAAGT (an inverted repeat of the first sequence)  
are recognized by the Cre recombinase and are the recognition sequences in Cre  
recombinase sites. Variant Cre recombinases recognize sites having different  
10 recognition sequences. The non-specific sequences (positions with "N" in the  
5 recognition sequence), although not limited in sequence, must be compatible  
with the non-specific sequences of the partner recombination site in order for  
recombination to be efficient. Thus, the non-specific sequences are the  
15 compatibility sequences of a recombinase site.

#### 15 C. Recombination Constructs

20 Recombination constructs are designed to provide an observable change  
when recombination between recombination sites occurs. Preferred  
20 recombination constructs include two pairs of recombination sites, one pair  
having a variant sequence and another pair having a sequence recognized by  
non-mutant recombinase (for example, wild type recombinase). Sites in the  
25 first pair are referred to as variant recombination sites. Generally,  
25 recombination constructs include a first nucleic acid sequence that includes a  
first reporter gene and first and second recombination sites, where the first and  
second recombination sites are variant recombination sites, and a second nucleic  
30 acid sequence that includes a second reporter gene and third and fourth  
20 recombination sites, where the third and fourth recombination sites can be  
recombined by a non-mutant recombinase. The first and second nucleic acid  
35 sequences need not be present on the same vector or on the same nucleic acid  
molecule (for example, the chromosome), although this is preferred. It is  
preferred that recombination constructs be embodied in vectors, such as  
25 plasmids.

40 In one embodiment of the disclosed recombination constructs, the  
sequence of the recombination sites in the constructs are chosen such that the  
recognition sequences of the first and second recombination sites differ from the  
recognition sequences of the third and fourth recombination sites. The sequence  
45 of the recombination sites can also be chosen such that the compatibility  
30 sequences of the first and second recombination sites differ from the  
compatibility sequences of the third and fourth recombination sites such that the

5 first and second recombination sites cannot recombine with the third and fourth  
recombination sites. The sequence of the recombination sites can also be  
chosen such that the compatibility sequences of the first and second  
10 recombination sites are sufficiently similar to allow recombination between the  
5 first and second recombination sites, and such that the compatibility sequences  
of the third and fourth recombination sites are sufficiently similar to allow  
recombination between the third and fourth recombination sites. The above  
15 sequence relationships result in constructs where the first and second  
recombination sites can recombine (in the presence of a recombinase that  
10 recognizes the sites), the third and fourth recombination sites can recombine,  
but where the neither the first nor second recombination site can recombine with  
20 either the third or fourth recombination site (since differences in the  
compatibility sequences prevent recombination).

Arriving at recombination sites having relationships as described above  
25 is preferably accomplished in the following way. Starting with a given  
recombination site sequence (which can be recombined by a non-mutant  
recombinase), parallel changes are made in the compatibility sequences of the  
first and second recombination sites. These altered recombination sites should  
30 then be tested to make sure that the non-mutant recombinase can still mediate  
20 their recombination. This helps insure that compatibility sequence changes  
have not inadvertently affected the function of the recombination sites. Once  
this is confirmed, changes can be made to the recognition sequences of the first  
35 and second recombination sites. These changes result in variant recombination  
sites for which variant recombinases can be identified using the method  
25 disclosed herein. The resulting recombination sites have the desired properties:  
40 incompatibility with the third and fourth recombination sites and variant  
recognition sequences that extend the range of recombination-competent sites.

The recombination sites can have a variety of properties and  
relationships that make them useful for particular purposes. For example, the  
45 recombination sites can be designed such that the first and second  
30 recombination sites cannot be recombined by non-mutant recombinase to a  
significant extent. This allows separate assessment of cleavage by mutant and  
50

5 non-mutant recombinase. It is also useful if the first and second recombination sites have identical sequences, and the third and fourth recombination sites have identical sequences.

10 5 Recombination between the recombination sites can have a variety of effects that allows detection of recombination. For example, the constructs can be designed such that recombination between the first and second recombination sites alters the expression of the first reporter gene, where recombination between the first and second recombination sites is determined by determining if expression of the first reporter gene is altered; recombination 15 between the third and fourth recombination sites alters the expression of the second reporter gene, where recombination between the third and fourth recombination sites is determined by determining if expression of the second reporter gene is altered; recombination between the first and second recombination sites allows the first reporter gene to be expressed; the first 20 nucleic acid sequence includes a spacer sequence flanked by the first and second recombination sites, where the spacer sequence interrupts the first reporter gene such that the first reporter gene is not expressed, and where recombination of the first and second recombination sites excises the spacer sequence which allows the first reporter gene to be expressed; and/or a portion 25 of the first reporter gene is inverted, wherein the inverted portion of the first reporter gene is flanked by the first and second recombination sites, wherein recombination of the first and second recombination sites inverts the inverted portion of the first reporter gene which allows the first reporter gene to be expressed.

30 25 The constructs can also be designed such that recombination between the first and second recombination sites prevents expression of the first reporter gene; the first reporter gene is flanked by the first and second recombination sites, where recombination of the first and second recombination sites excises the first reporter gene which prevents expression of the first reporter gene; a 45 30 portion of the first reporter gene is flanked by the first and second recombination sites, where recombination of the first and second recombination sites inverts the flanked portion of the first reporter gene which prevents

5 expression of the first reporter gene: recombination between the third and fourth  
recombination sites allows the second reporter gene to be expressed; and/or the  
second nucleic acid sequence includes a spacer sequence flanked by the third  
10 and fourth recombination sites, where the spacer sequence interrupts the second  
5 reporter gene such that the second reporter gene is not expressed, and where  
recombination of the third and fourth recombination sites excises the spacer  
sequence which allows the second reporter gene to be expressed.

15 The constructs can also be designed such that a portion of the second  
reporter gene is inverted, where the inverted portion of the second reporter gene  
10 is flanked by the third and fourth recombination sites, and where recombination  
of the third and fourth recombination sites inverts the inverted portion of the  
second reporter gene which allows the second reporter gene to be expressed;  
20 recombination between the third and fourth recombination sites prevents  
expression of the second reporter gene to be expressed; the second reporter gene  
15 is flanked by the third and fourth recombination sites, where recombination of  
the third and fourth recombination sites excises the second reporter gene which  
prevents expression of the second reporter gene; and/or a portion of the second  
reporter gene is flanked by the third and fourth recombination sites, where  
30 recombination of the third and fourth recombination sites inverts the flanked  
20 portion of the second reporter gene which prevents expression of the second  
reporter gene.

35 Expression of a reporter gene can include transcription of the gene,  
translation of the transcript (if the gene encodes a protein), and/or production of  
an active protein. As used herein, whether a reporter gene is expressed depends  
25 on the context. In general, a gene is considered to be expressed if it produces  
the expression product to be detected. Such expression products include full or  
partial transcripts of the gene, full or partial proteins, including active or  
inactive forms of the proteins, translated from the transcript. Since the goal in  
using reporter genes in the disclosed method is the detection of expression, any  
45 of these forms of expression product can be the object of detection. For  
30 example, if the gene's transcript is to be detected, the gene will be considered to  
be expressed if it produces the transcript, regardless of whether the transcript is

5 translated or whether the resulting protein is active. If the an active protein  
encoded by the gene is to be detected, the gene is not expressed unless active  
protein is produced--mere transcription of the gene, or even translation to  
10 produce an inactive protein, will not be enough in this context. As a  
5 consequence, the expression product to be detected will influence the manner in  
which reporter genes should be interrupted or invented in the disclosed  
constructs. For example, nearly any interruption of a reporter gene would  
15 prevent expression of an active protein encoded by the gene. On the other hand,  
an interruption of the coding region will usually not prevent production of a  
10 transcript. The structure of the disclosed constructs should be designed with  
these principles in mind. As used herein, an inactive expression product refers  
20 to an expression product that does not have an activity exhibited by the active  
form of the expression product where that activity is required for detection of  
expression in the assay scheme being used.

15 The constructs can be designed such that the first nucleic acid sequence  
is a first nucleic acid construct and the second nucleic acid sequence is on a  
second nucleic acid construct; the first nucleic acid construct is an  
extrachromosomal vector and the second nucleic acid construct is in the genome  
30 of a host cell; and/or the first and second nucleic acid constructs are on the same  
20 nucleic acid construct.

#### D. Reporter Genes

35 Reporter genes are used to monitor whether recombination occurs in the  
disclosed constructs. Reporter genes can be any gene the expression of which  
can be detected either directly or indirectly. These include genes encoding  
25 enzymes, such as  $\beta$ -galactosidase, luciferase, and alkaline phosphatase, that can  
produce specific detectable products, and genes encoding proteins that can be  
40 directly detected. Virtually any protein can be directly detected by using, for  
example, specific antibodies to the protein. A preferred reporter protein that can  
be directly detected is the green fluorescent protein (GFP). GFP, from the  
45 jellyfish *Aequorea victoria*, produces fluorescence upon exposure to ultraviolet  
30 light without the addition of a substrate (Chalfie *et al.*, *Science* 263:802-5  
(1994)). A number of modified GFPs have been created that generate as much

5 as 50-fold greater fluorescence than does wild type GFP under standard conditions (Cormack *et al.*, *Gene* 173:33-8 (1996); Zolotukhin *et al.*, *J. Virol* 70:4646-54 (1996)). This level of fluorescence allows the detection of low levels of expression in cells.

10 5 Reporter genes encoding proteins producing a fluorescent signal are useful since such a signal allows cells to be sorted using FACS. Another way of sorting cells based on expression of the reporter gene involves using the reporter protein as a hook to bind cells. For example, a cell surface protein such as a receptor protein can be bound by a specific antibody. Cells expressing such a  
15 protein can be captured by, for example, using antibodies bound to a solid substrate, using antibodies bound to magnetic beads, or capturing antibodies bound to the reporter protein. Many techniques for the use of antibodies as capture agents are known and can be used with the disclosed method.

20 The reporter gene can also encode an expression product that regulates the expression of another gene. This allows detection of expression of the reporter gene by detecting expression of the regulated gene. For example, a repressor protein can be encoded by the reporter gene. Loss of expression of the reporter gene (via recombination) would then result in derepression of the regulated gene. This type of indirect detection allows positive detection of loss  
25 of the expression of the reporter gene by the effector RNA molecule. One preferred form of this type of regulation is the use of an antibiotic resistance gene regulated by a repressor protein encoded by the reporter gene. By exposing the host cells to the antibiotic, only those cells in which expression of the reporter gene has been inhibited will grow since expression of the antibiotic resistance gene will be derepressed.

#### 40 E. Expression Sequences

The reporter genes can be expressed using any suitable expression sequences. Numerous expression sequences are known and can be used for expression of the reporter genes. Expression sequences can generally be  
45 classified as promoters, terminators, and, for use in eukaryotic cells, enhancers. Expression in prokaryotic cells also requires a Shine-Dalgarno sequence just upstream of the coding region for proper translation initiation. Inducible

5 promoters are preferred for use with the first reporter gene since it is preferred that expression of the first reporter gene be adjustable.

Promoters suitable for use with prokaryotic hosts illustratively include the  $\beta$ -lactamase and lactose promoter systems, tetracycline (tet) promoter, 10 alkaline phosphatase promoter, the tryptophan (trp) promoter system and hybrid promoters such as the tac promoter. However, many other functional bacterial promoters are suitable. Their nucleotide sequences are generally known.

Suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase, enolase, glyceraldehyde-3-phosphate 15 dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triphosphate isomerase, phosphoglucose isomerase, and glucokinase.

Examples of inducible yeast promoters suitable for use in the disclosed vectors include the promoter regions for alcohol dehydrogenase 2, isocytichrome C, 20 acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Yeast enhancers also are advantageously used with yeast promoters.

Preferred promoters for use in mammalian host cells include promoters 25 from polymoma virus, Simian Virus 40 (SV40), adenovirus, retroviruses, hepatitis B virus, herpes simplex virus (HSV), Rous sarcoma virus (RSV), mouse mammary tumor virus (MMTV), and most preferably cytomegalovirus (CMV), or from heterologous mammalian promoters such as the  $\beta$  actin promoter. Particularly preferred are the early and late promoters of the SV40 30 virus and the immediate early promoter of the human cytomegalovirus, MMTV LTR, RSV-LTR, and the HSV thymidine kinase promoter.

Transcription of the reporter gene by higher eukaryotes can be increased by inserting an enhancer sequence into the vector. Many enhancer 35 sequences are now known from mammalian genes (globin, elastase, albumin, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the

5 replication origin, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

10 The disclosed vectors preferably also contain sequences necessary for accurate 3' end formation of both reporter and effector RNAs. In eukaryotic cells, this would be a polyadenylation signal. In prokaryotic cells, this would be a transcription terminator.

#### Method

##### A. Identification of Variant Recombinases

15 The disclosed method involves producing mutant recombinases and testing the mutant recombinases with specially designed constructs. The constructs contain variant recombination sites that are not recognized by non-mutant recombinase but will undergo recombination in the presence of a mutant recombinase with altered specificity. The disclosed method also involves determining whether a variant recombinase retains the ability to mediate recombination at recombination sites recognized by non-variant recombinase. This can be accomplished by using constructs containing recombination sites recognized by non-variant recombinase. Recombination at these recombination sites can be monitored or detected by any suitable means. It is preferred that recombination is detected by screening or selection based on the expression or lack of expression of a reporter gene. This can be accomplished by using constructs containing a reporter gene associated with the recombination sites such that the reporter gene is rearranged or deleted, or a spacer sequence interrupting the reporter gene is rearranged or deleted, as a result of recombination at the recombination sites. Recombination of such constructs will result in a loss of expression of the reporter gene, where the construct contained a functional reporter gene, or in a gain in expression of the reporter gene, where the construct contained a non-functional reporter gene.

##### 1. Production of Mutant Recombinases

45 30 Mutant recombinases can be produced by any suitable technique. In general, all that is required is a method of generating a variety of recombinase proteins having a variety of amino acid sequences. the most preferred way of doing this is to mutagenize or alter nucleic acid encoding the recombinase and

5 then expressing the mutant recombinases. Numerous techniques for introducing  
alterations into nucleic acid sequences are known and can be used in the  
disclosed method. For example, alterations can be made by chemical  
10 mutagenesis, introduction of degenerate nucleic acid fragments into the base  
5 nucleic acid molecule, and low fidelity PCR. The goal of this mutagenesis or  
alteration will be the generation of a population or set of mutant recombinases  
having a variety of sequences. The broader the range of variants, the more raw  
15 material for the identification process.

## 2. Identification of Variant Recombinases That Recognize Variant 10 Recombination Sites

Variant recombinases that can mediate recombination at variant  
20 recombination sites are identified in the disclosed method by selecting for,  
screening for, or otherwise detecting recombination of specially designed  
constructs having variant recombination sites. Recombination at variant  
15 recombination sites can be monitored or detected by any suitable means. It is  
25 preferred that recombination is detected by screening or selection based on the  
expression or lack of expression of a reporter gene. This can be accomplished  
by using constructs containing a reporter gene associated with the variant  
30 recombination sites such that the reporter gene is rearranged or deleted, or a  
20 spacer sequence interrupting the reporter gene is rearranged or deleted, as a  
result of recombination at the recombination sites. Recombination of such  
constructs will result in a loss of expression of the reporter gene, where the  
35 construct contained a functional reporter gene, or in a gain in expression of the  
reporter gene, where the construct contained a non-functional reporter gene.

## 3. Identification of Variant Recombinases That Recognize Non- 40 Variant Recombination Sites

Variant recombinases that can mediate recombination at recombination  
sites recognized by non-variant recombinase (non-variant recombination sites)  
are identified in the disclosed method by selecting for, screening for, or  
45 30 otherwise detecting recombination of specially designed constructs having  
recombination sites recognized by non-variant recombinase. Recombination at  
these recombination sites can be monitored or detected by any suitable means.

5 It is preferred that recombination is detected by screening or selection based on  
the expression or lack of expression of a reporter gene. This can be  
accomplished by using constructs containing a reporter gene associated with the  
10 recombination sites recognized by non-variant recombinase such that the  
5 reporter gene is rearranged or deleted, or a spacer sequence interrupting the  
reporter gene is rearranged or deleted, as a result of recombination at the  
recombination sites. Recombination of such constructs will result in a loss of  
15 expression of the reporter gene, where the construct contained a functional  
reporter gene, or in a gain in expression of the reporter gene, where the  
20 construct contained a non-functional reporter gene.

It is preferred that the ability of a variant recombinase to mediate  
20 recombination at both variant recombination sites and recombination sites  
recognized by non-variant recombinase be assessed in the same system (such as  
a cell strain) either sequentially or simultaneously. When variant recombinases  
15 are tested for activity on both variant recombination sites and recombination  
25 sites recognized by non-variant recombinase in the same system or at the same  
time, it is preferred that two different reporter genes which can be separately  
detected or monitored be used. In this case, a first reporter gene can be  
30 associated with the variant recombination sites and a second reporter gene can  
20 be associated with recombination sites recognized by non-variant recombinase.

#### **B. Use of Variant Recombinases**

Variant recombinases produced in the disclosed method can be used for  
35 any purpose that unmodified recombinases can be used. The advantage is that  
the variant recombinases have a different or broader site specificity. In general,  
25 the disclosed variant recombinases can be used to mediate recombination of any  
40 nucleic acid in any setting, including *in vitro*, in cell culture, and *in vivo*.  
Recombination can be obtained in single celled organisms, such as bacterial  
cells, fungal cells, yeast cells, prokaryotic cells, and archae bacterial cells, the  
45 cells of multicellular organisms, including plants and animals, both in the  
30 organism and in culture. The disclosed variant recombinases can also be used  
in combination with other recombinases (including other variant recombinases)  
having a different site specificity. Such combinations allow more complex

5 recombination schemes to be used. Examples of such schemes are discussed below.

10 For some uses of the disclosed recombinases, first, second, and fourth DNA sequences comprising a first recombination site, a second recombination site, and a third recombination site, respectively, are introduced into cells. As used herein the expression "recombination site" means a nucleotide sequence at which a recombinase or variant recombinase can catalyze a site-specific recombination.

15 Methods for introducing a DNA sequence into cells are known in the art. 20 These methods typically include the use of a DNA vector to introduce the sequence into the DNA of a single or limited number of eukaryotic cells and then growing such cell or cells to generate a suitable population of cells. As used herein, the term "vector" includes plasmids, viruses, and viral vectors. Preferably, the DNA sequences are introduced by a plasmid capable of 25 transforming a selected cell while carrying a DNA sequence. The particular vector which is employed to introduce the DNA sequence into a selected cell is not critical.

30 In the present method, the recombination sites are contacted with a variant recombinase, thereby producing the site specific recombination. A preferred means of contacting the DNA to be recombined with a variant recombinase is to place the DNA to be recombined into a cell expressing 35 nucleic acid encoding the variant recombinase. Preferably, expression of the variant recombinase is under the control of a regulatory nucleotide sequence. Suitable regulatory nucleotide sequences are known in the art. The regulatory 40 nucleotide sequence which is employed with a selected eukaryotic cell is not critical to the method. A partial list of suitable regulatory nucleotide sequences includes the long terminal repeat of Moloney sarcoma virus described by Blochlinger and Diggelmann, *Mol. Cell Bio.*, 4:2929-2931 (1984); the mouse metallothionein-I promoter described by Pavlakis and Hamer, *Proc. Natl. Acad. Sci USA*, 80:397-401 (1983); the long terminal repeat of Rous sarcoma virus 45 described by Gorman et al., *Proc. Natl. Acad. Sci USA*, 79:6777-6781 (1982);

5 and the early region promoter of SV40 described by Southern and Berg, *J. Mol. Appl. Genet.*, 1:327-341 (1982).

10 In an embodiment where the cells are yeast, suitable regulatory nucleotide sequences include GAL1, GAL10, ADH1, CYC1, and TRP5  
5 promoters. GAL1 and GAL10 promoters are present on plasmid pBM150 which is described by Johnston and Davis, *Molec. Cell. Biol.*, 4:1440 (1984). The  
15 ADH1 promoter, also called ADC1, is present on plasmid pAAH5 which is described by Ammer, *Methods Enzymol.*, 101:192 (1983). The CYC1 promoter is described by Stiles et al., *Cell*, 25:277 (1981). The TRP5 promoter is  
10 described by Zalkin and Yanofsky, *J. Biol. Chem.*, 257:1491 (1982). Preferably, the regulatory nucleotide sequence is a GAL1 promoter.

20 In one embodiment where the cell is yeast, the first, second, and optionally, third and fourth DNA sequences are introduced into one strain of yeast. Alternatively, the DNA sequences are introduced into two different  
15 strains of yeast of opposite mating types which are subsequently mated to form a single strain having all three or four DNA sequences. Preferably, the plasmid contains either (1) a nucleotide sequence of DNA homologous to a resident  
25 yeast sequence to permit integration into the yeast DNA by the yeast's recombination system or (2) a nucleotide sequence of DNA which permits  
30 autonomous replication in yeast. One nucleotide sequence which permits autonomous replication in yeast is an ARS sequence described by Stinchcomb et al., *Nature*, 282:39 (1979). A partial list of plasmids capable of transforming  
35 yeast includes YIP5, YRP17 and YEP24. These plasmids are disclosed and described by Botstein and Davis, *The Molecular Biology of the Yeast Saccharomyces*, Metabolism and Gene Expression (ed. Strathern et al.), (Cold  
40 Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), at page 607.

45 Since most recombination sites are asymmetrical nucleotide sequences, two recombination sites on the same DNA molecule can have the same or opposite orientations with respect to each other. Recombinations between  
30 recombination sites in the same orientation result in a deletion of the DNA segment located between the two recombination sites and a connection between the resulting ends of the original DNA molecule. The deleted DNA segment

5 forms a circular molecule of DNA. The original DNA molecule and the  
resulting circular molecule each contain a single recombination site (see Figure  
2). Recombination between recombination sites in opposite orientations on the  
10 same DNA molecule result in an inversion of the nucleotide sequence of the  
5 DNA segment located between the two recombination sites (see Figure 2). In  
addition, reciprocal exchange of DNA segments proximate to recombination  
sites located on two different DNA molecules can occur. All of these  
15 recombination events are catalyzed by recombinases, including the disclosed  
variants and wild type recombinases.

10 Recombination using the disclosed variant recombinases can be used *in*  
*vitro* to produce site-specific recombination of nucleic acid molecules. This is  
20 useful for a wide variety of manipulations that currently employ wild type  
recombinases or involve traditional restriction enzyme cleavage followed by  
ligation. Examples include recombination of libraries of DNA fragments into  
15 vectors or in desired structures, and labeling of DNA via recombination.  
Recombined DNA formed by *in vitro* recombination can then be introduced into  
cells. For example, constructs formed *in vitro* can be introduced into cells to  
25 resolve the structures formed *in vitro* or to select active structures. In particular,  
large concatemers of subject DNA and spacer/vector fragments can be made,  
30 introduced into cells, and circularized into vector units in the cells. Such  
20 recombination could also be performed *in vitro* if desired.

35 The disclosed variant recombinases can be used to label DNA by  
recombining a DNA molecule of interest with a labeled DNA molecule. Use of  
a recombinase for labeling is advantageous since it involves fewer steps than  
25 traditional labeling via DNA synthesis or ligation. These considerations are  
particularly important when large DNA molecules (over 20 kb) are to be labeled  
40 since such large molecules will fragment more the more they are manipulated.

45 Recombination mediated by the disclosed variant recombinases and  
variant recombination sites can be used to manipulate a host cell genome as  
30 desired and simultaneously introduce a marker gene flanked by the recognition  
sites of a second recombinase. After selection, leading to an accumulation of  
cells carrying the desired genomic alteration, one could simply remove the

5 marker gene by expression of the second site-specific recombinase. A large  
number of recombinases suitable for this purpose exists in nature, including  $\lambda$   
Integrase (Int), yeast Fip, etc. (Nunes-Döby et al., *Nucl. Acids Res.*, 26:391-406  
10 (1998)). Variant recombinases having different site specificity can also be used.

5 Since the disclosed variant recombinases recognize both wild type  
recombination sites and variant recombination sites that are not recognized by  
wild type recombinase, wild type recombinase and variant recombinases can be  
used to mediate sequential recombination between nucleic acids containing a  
15 combination of wild type recombination sites and variant recombination sites.

10 For example, generation of knockout animals and plants can be made more  
efficient by using a structure wild type recombination site-selectable marker-  
wild type recombination site-endogenous gene-variant recombination site  
(rather than the conventional wild type recombination site-selectable marker-  
wild type recombination site-endogenous gene-wild type recombination site).

15 Such a structure allows the selectable marker to be removed by the action of  
wild type recombinase without disturbing the gene since wild type recombinase  
will not recognize the variant recombination site to any significant degree. The  
endogenous gene can then be removed later by the action of a variant  
20 recombinase since the disclosed variant recombinases recognize both wild type  
and variant recombination sites.

In a preferred embodiment of the disclosed method, the first and second  
DNA sequences are introduced into cells connected by a pre-selected DNA  
35 segment. The segment can be a gene or any other sequence of  
deoxyribonucleotides of homologous, heterologous or synthetic origin.

25 Preferably, the pre-selected DNA segment is a gene for a structural protein, an  
enzyme, or a regulatory molecule. If the first and second recombination sites  
have the same orientation, activation of the regulatory nucleotide sequence  
produces a deletion of the pre-selected DNA segment. If the first and second  
40 recombination sites have opposite orientation, activation of the regulatory  
nucleotide sequence produces an inversion of the nucleotide sequence of the  
pre-selected DNA segment.

5 If a fourth DNA sequence (containing the third recombination site) is  
also introduced into cells, it is preferred that the second and fourth DNA  
sequences be introduced into cells connected by a second pre-selected DNA  
10 segment. The second segment can be a gene or any other sequence of  
5 deoxyribonucleotides of homologous, heterologous or synthetic origin.  
Preferably, the second pre-selected DNA segment is a gene for a structural  
protein, an enzyme, or a regulatory molecule. If the second and third  
15 recombination sites have the same orientation, activation of the regulatory  
nucleotide sequence produces a deletion of the second pre-selected DNA  
10 segment. If the second and third recombination sites have opposite orientation,  
activation of the regulatory nucleotide sequence produces an inversion of the  
20 nucleotide sequence of the second pre-selected DNA segment.

Combinations of wild type and variant recombination sites, and  
combinations of different orientations of the recombination sites, in DNA  
25 introduced into cells can multiply recombination options. For example, if the  
first and second recombination sites are wild type recombination sites and the  
third recombination site is a variant recombination site (all in the same  
orientation) then wild type recombinase can produce a deletion of the first pre-  
30 selected DNA segment (but not the second) and a variant recombinase can  
20 produce a deletion of the first, second, or both pre-selected DNA segments.  
This arrangement allows sequential deletion of the first and second pre-selected  
DNA segments.

35 If the first and second recombination sites are wild type recombination  
sites and the third recombination site is a variant recombination site, and the  
25 first recombination site has the opposite orientation from the second and third  
recombination sites (which, of course, have the same orientation) then wild type  
40 recombinase can produce an inversion of the first pre-selected DNA segment  
and a variant recombinase can produce a deletion of the second pre-selected  
DNA segment (and/or produce an inversion of the first pre-selected DNA  
45 segment or the entire section spanning the first, second, and third recombination  
30 sites).

5 If the first and third recombination sites are wild type recombination sites and the second recombination site is a variant recombination site, and the second recombination site has the opposite orientation from the first and third recombination sites (which, of course, have the same orientation) then wild type  
10 5 recombinase can produce a deletion of the entire section spanning the first, second, and third recombination sites, and a Cre variant can produce an inversion of the first, second, or both pre-selected DNA segments.

15 If the first and third recombination sites are wild type recombination sites and the second recombination site is a variant recombination site, and the first recombination site has the opposite orientation from the second and third recombination sites (which, of course, have the same orientation) then wild type  
20 recombinase can produce an inversion of the entire section spanning the first, second, and third recombination sites, and a variant recombinase can produce a deletion of the second pre-selected DNA segments and an inversion of the first  
25 pre-selected DNA segment.

Many more combinations of wild type and variant recombination sites and or recombination site orientations are possible. For example, the variant recombinase can also be used with a different variant recombinase having a different site specificity rather than wild type recombinase. The above  
30 20 examples illustrate the general principles involved in designing specific recombinations that may be desired. It should be understood that the above combinations of recombination sites can be extended to the use of more recombination sites (that is more than three) and more intervening, pre-selected DNA segments.

25 For some uses of the disclosed recombinases, first, second, and fourth DNA sequences comprising a first lox site, a second lox site, and a third lox site, respectively, are introduced into cells. As used herein the expression "lox site" means a nucleotide sequence at which the gene product of the cre gene, referred to herein as Cre, and/or the disclosed Cre variants, can catalyze a site-specific recombination. LoxP site is a 34 base pair nucleotide sequence (Figure  
40 30 1) which can be directly synthesized or isolated from bacteriophage P1 by methods known in the art. The Lox P site is an example of a wild type lox site.

5 One method for isolating a LoxP site from bacteriophage P1 is disclosed by  
Hoess et al., *Proc. Natl. Acad. Sci. USA*, 79:3398-3402 (1982). The LoxP site  
consists of two 13 base pair inverted repeats separated by an 8 base pair spacer  
10 region. The nucleotide sequences of the insert repeats and the spacer region are  
as follows.

ATAACTTCGTATAATGTATGCTATACGAAGTTAT

Other wild type lox sites include LoxB, LoxL and LoxR sites which are  
nucleotide sequences isolated from *E. coli*. These sequences are disclosed and  
described by Hoess et al., *Proc. Natl. Acad. Sci. USA*, 79:3398-3402 (1982).

15 Preferred wild type lox sites are LoxP or LoxC2. Lox sites can also be produced  
by a variety of synthetic techniques which are known in the art. For example,  
synthetic techniques for producing lox sites are disclosed by Ito et al., *Nuc. Acid*  
20 *Res.*, 10:1755 (1982) and Ogilvie et al., *Science*, 214:270 (1981).

The gene product of the cre gene is a recombinase herein designated  
25 "Cre" which effects site-specific recombination of DNA at lox sites. As used  
herein, the expression "cre gene" means a nucleotide sequence which codes for  
a gene product which effects site-specific recombination of DNA in cells at lox  
sites. One cre gene (the wild type cre gene) can be isolated from bacteriophage  
30 P1 by methods known in the art. One method for isolating a cre gene from  
bacteriophage P1 is disclosed by Abremski et al., *Cell*, 32:1301-1311 (1983).

Genes engineered into cells for producing a foreign protein are often  
placed under the control of a highly active promoter. The activity of the  
35 promoter can result in an overproduction of the protein which interferes with the  
growth of the engineered cell. This overproduction of the protein can make it  
difficult to grow the engineered cell in sufficient quantity to make protein  
25 production economically feasible. The present invention provides a method  
whereby engineered cells can be grown to a desired density prior to expressing  
the engineered gene. The engineered gene is expressed, as desired, by  
40 activating a regulatory nucleotide sequence responsible for controlling  
expression of DNA encoding a variant recombinase. Methods of controlling the  
45 expression of an engineered gene include the following:

(1) A DNA segment flanked by recombination sites in the same orientation is introduced into DNA in a cell between a promoter and an engineered gene to render the promoter incapable of expressing the gene. A second DNA sequence comprising a regulatory nucleotide sequence and DNA encoding a variant recombinase is also introduced in the DNA. After the engineered cells are grown to a desired density, the regulatory nucleotide sequence is activated thereby effecting expression of the variant recombinase and producing a deletion of the DNA segment. The engineered gene would then be expressed.

(2) A gene for a regulatory molecule flanked by recombination sites in the same orientation is introduced into DNA in a cell. The regulatory molecule inhibits expression of an engineered gene. A second DNA sequence comprising a regulatory nucleotide sequence and DNA encoding a variant recombinase is also introduced into the DNA. After the engineered cells are grown to a desired density, the regulatory nucleotide sequence is activated thereby effecting expression of the variant recombinase and producing a deletion of the gene for the regulatory molecule. The engineered gene would then be expressed.

(3) An engineered gene lacking a promoter and flanked by two recombination sites in opposite orientations is introduced into DNA in a cell such that the 3' end of the gene lies adjacent to the transcription start site of a regulatory nucleotide sequence. A second DNA sequence comprising a regulatory nucleotide sequence and DNA encoding a variant recombinase is also introduced into the DNA. Since the engineered gene would be transcribed in the antisense direction, no engineered protein would be produced. After the engineered cell is grown to a desired density, the regulatory nucleotide sequence is activated thereby effecting expression of the variant recombinase and producing an inversion of the desired gene. The engineered gene could then be transcribed in the proper direction and expressed.

Numerous methods and techniques have been developed for the use of Cre recombinase and other, similar recombinases such as FLP. The disclosed variant recombinases can also be used in any of these methods. Adaptation of these methods to the use of the disclosed variant recombinases is

5 straightforward. Generally, all that is required is substitution of a variant  
recombinase (or a gene expressing a variant recombinase) for the original  
recombinase (or recombinase gene) and, if appropriate, substitution of variant  
10 (or wild type) recombination sites for the original recombination sites used in  
5 the method.

Examples of methods involving wild type recombinases and wild type  
recombination sites that can be adapted for use with the disclosed variant  
15 recombinases and recombination sites include recombination of DNA in phage  
packaging systems, recombination of DNA to form phage display libraries (for  
20 example, Fisch *et al.*, *Proc Natl Acad Sci USA* 93(15):7761-6 (1996), and  
Waterhouse *et al.*, *Nucleic Acids Res* 21(9):2265-6 (1993), and other uses (for  
example, (Sauer *et al.*, *Proc. Natl. Acad. Sci. USA* 84: 9108-9112 (1987),  
Mullins *et al.*, *Nucleic Acids Res* 25(12):2539-40 (1997), Aoki *et al.*, *Mol Med*  
5(4):224-31 (1999)).

15 Other examples of specific methods in which the disclosed variant  
recombinases can be used or substituted include methods disclosed in U.S.  
Patent No. 5,888,981, U.S. Patent No. 5,888,732, U.S. Patent No. 5,885,836,  
U.S. Patent No. 5,885,793, U.S. Patent No. 5,885,779, U.S. Patent No.  
30 5,885,776, U.S. Patent No. 5,882,893, U.S. Patent No. 5,882,888, U.S. Patent  
20 No. 5,877,400, U.S. Patent No. 5,871,907, U.S. Patent No. 5,866,755, U.S.  
Patent No. 5,866,361, U.S. Patent No. 5,859,310, U.S. Patent No. 5,858,657,  
U.S. Patent No. 5,854,067, U.S. Patent No. 5,851,808, U.S. Patent No.  
35 5,849,995, U.S. Patent No. 5,849,989, U.S. Patent No. 5,849,708, U.S. Patent  
No. 5,849,572, U.S. Patent No. 5,849,571, U.S. Patent No. 5,849,553, U.S.  
25 Patent No. 5,844,079, U.S. Patent No. 5,843,744, U.S. Patent No. 5,843,742,  
U.S. Patent No. 5,843,694, U.S. Patent No. 5,840,540, U.S. Patent No.  
40 5,837,844, U.S. Patent No. 5,837,242, U.S. Patent No. 5,834,202, U.S. Patent  
No. 5,830,729, U.S. Patent No. 5,830,698, U.S. Patent No. 5,830,461, U.S.  
Patent No. 5,817,492, U.S. Patent No. 5,814,618, U.S. Patent No. 5,814,500,  
45 30 U.S. Patent No. 5,814,300, U.S. Patent No. 5,807,995, U.S. Patent No.  
5,807,708, U.S. Patent No. 5,801,030, U.S. Patent No. 5,800,998, U.S. Patent  
No. 5,795,734, U.S. Patent No. 5,795,726, U.S. Patent No. 5,792,833, U.S.

5 Patent No. 5,792,632, U.S. Patent No. 5,789,156, U.S. Patent No. 5,777,194,  
U.S. Patent No. 5,776,449, U.S. Patent No. 5,773,697, U.S. Patent No.  
5,770,384, U.S. Patent No. 5,767,376, U.S. Patent No. 5,763,240, U.S. Patent  
10 No. 5,756,671, U.S. Patent No. 5,744,343, U.S. Patent No. 5,744,336, U.S.  
5 Patent No. 5,736,377, U.S. Patent No. 5,733,744, U.S. Patent No. 5,733,743,  
U.S. Patent No. 5,733,733, U.S. Patent No. 5,731,182, U.S. Patent No.  
5,723,765, U.S. Patent No. 5,723,333, U.S. Patent No. 5,723,287, U.S. Patent  
15 No. 5,721,367, U.S. Patent No. 5,721,118, U.S. Patent No. 5,700,470, U.S.  
Patent No. 5,686,595, U.S. Patent No. 5,679,523, U.S. Patent No. 5,677,177,  
20 U.S. Patent No. 5,658,772, U.S. Patent No. 5,656,438, U.S. Patent No.  
5,654,182, U.S. Patent No. 5,654,168, U.S. Patent No. 5,650,491, U.S. Patent  
No. 5,650,308, U.S. Patent No. 5,650,298, U.S. Patent No. 5,643,727, U.S.  
Patent No. 5,641,866, U.S. Patent No. 5,641,748, U.S. Patent No. 5,639,726,  
25 U.S. Patent No. 5,635,381, U.S. Patent No. 5,629,179, U.S. Patent No.  
15 5,629,159, U.S. Patent No. 5,614,389, U.S. Patent No. 5,612,205, U.S. Patent  
No. 5,596,089, U.S. Patent No. 5,591,609, U.S. Patent No. 5,589,362, U.S.  
Patent No. 5,539,094, U.S. Patent No. 5,530,191, U.S. Patent No. 5,527,695,  
30 U.S. Patent No. 5,510,099, U.S. Patent No. 5,478,731, U.S. Patent No.  
5,441,884, U.S. Patent No. 5,434,066, U.S. Patent No. 5,378,618, U.S. Patent  
20 No. 5,354,668, U.S. Patent No. 5,334,515, U.S. Patent No. 5,300,431, and U.S.  
Patent No. 4,959,317.

#### 1. Use of Variants Recombinases in Plants and Plant Cells

35 Methods for introducing a DNA sequence into plant cells are known in  
the art. Nucleic acids can generally be introduced into plant protoplasts, with or  
25 without the aid of electroporation, polyethylene glycol, or other processes  
known to alter membrane permeability. Nucleic acid constructs can also be  
40 introduced into plants using vectors comprising part of the Ti- or Ri-plasmid, a  
plant virus, or an autonomously replicating sequence. Nucleic acid constructs  
can also be introduced into plants by microinjection or by high-velocity  
45 microprojectiles, also termed "particle bombardment" or "biolistics" (Sanford, J.  
C., *Tibtech* 6: 299 (1988)), directly into various plant parts. The preferred  
30 means of introducing a nucleic acid fragment into plant cells involves the use of

5 *A. tumefaciens* containing the nucleic acid fragment between T-DNA borders  
either on a disarmed Ti-plasmid (that is, a Ti-plasmid from which the genes for  
tumorigenicity have been deleted) or in a binary vector in trans to a disarmed  
10 Ti-plasmid. The *Agrobacterium* can be used to transform plants by inoculation  
5 of tissue explants, such as stems, roots, or leaf discs, by co-cultivation with  
plant protoplasts, or by inoculation of seeds or wounded plant parts.

Foreign genes can be introduced into a wide range of crop species.  
15 Thus, the disclosed variant recombinases and method are applicable to a broad  
range of agronomically or horticulturally useful plants. The particular method  
20 which is employed to introduce the DNA sequence into a selected plant cell is  
not critical. In a preferred embodiment, DNA sequences are introduced into  
plant cells by co-cultivation of leaf discs with *A. tumefaciens* essentially as  
described by Horsch et al., *Science*, 227: 1229-1231 (1985) omitting the nurse  
cultures.

15 In the present method, the recombination sites are contacted with a  
variant recombinase, thereby producing the site specific recombination. In one  
embodiment, a variant recombinase, or messenger RNA encoding a variant  
30 recombinase, is introduced into the cells directly by micro-injection, biolistics,  
or other protein or RNA introduction procedure. In a preferred embodiment,  
20 DNA encoding the variant recombinase is introduced into the plant cell under  
the control of a promoter that is active in plant cells. Suitable regulatory  
nucleotide sequences are known in the art. The promoter which is employed  
35 with a selected plant cell is not critical to the method of the invention. A partial  
list of suitable promoters include the 35S promoter of cauliflower mosaic virus  
25 described by Odell et al., *Nature*, 313: 810-812 (1985); the promoter from the  
nopaline synthase gene of *A. tumefaciens* described by Depicker et al., *J. of*  
40 *Mol. Appl. Genet.*, 1: 561-573 (1982); the promoter from a Rubisco small  
subunit gene described by Mazur and Chui, *Nucleic Acids Research* 13: 2373-  
2386 (1985); the 1' or 2' promoter from the TR-DNA of *A. tumefaciens*  
45 described by Velten et al., *EMBO J.* 12: 2723-2730 (1984); the promoter of a  
30 chlorophyll a/b binding protein gene described by Dunsmuir et al., *J. Mol. Appl.*  
*Genet.* 2: 285-300 (1983); the promoter of a soybean seed storage protein gene

described by Chen et al., *Proc. Natl. Acad. Sci. USA*, 83: 8560-8564 (1986); and the promoter from the wheat EM gene described by Marcotte et al., *Nature* 335: 454-457 (1988). Variant recombinases can be expressed throughout the plant generally in all cells at all stages of development, or expression of variant recombinases can be more specifically controlled through the use of promoters or regulatory nucleotide sequences having limited expression characteristics. Variant recombinases can be expressed in a tissue specific manner, for example only in roots, leaves, or certain flower parts. Variant recombinases can be expressed in a developmentally specific time period, for example only during seed formation or during reproductive cell formation. Expression of variant recombinases can also be placed under the control of a promoter that can be regulated by application of an inducer. In this case expression is off or very low until the external inducer is applied. Promoters active in plant cells have been described that are inducible by heat shock (Gurley et al., *Mol. Cell. Biol.* 6: 559-565 (1986)), ethylene (Bfogle et al., *Plant Cell* 1: 599-607 (1989)), auxin (Hagan and Guilfoyle, *Mol. Cell. Biol.* 5: 1197-1203 (1985)), abscisic acid (Marcotte et al., *Nature* 335: 454-457 (1988)), salicylic acid (EPO 332104A2 and EPO 337532A1), and substituted benzenesulfonamide safeners (WO 90/11361). Control of expression of variant recombinases by the safener-inducible promoter 2-2, or its derivatives, allows the expression to be turned on only when the inducing chemical is applied and not in response to environmental or phytohormonal stimuli. Thus expression can be initiated at any desired time in the plant life cycle. Preferably, the regulatory nucleotide sequence is a 35S promoter or a 2-2 promoter. The above techniques and materials can also be used to express wild type recombinase in plant cells if required by the particular recombination pattern to be accomplished.

One application of the disclosed variant recombinases is in controlling male fertility in a method for producing hybrid crops. Hybridization of a crop involves the crossing of two different lines to produce hybrid seed from which the crop plants are grown. Hybrid crops are superior in that more of the desired traits can be introduced into the production plants. For instance, quality traits such as oil content, herbicide resistance, disease resistance, adaptability to

5 environmental conditions, and the like, can be hybridized in offspring so that  
the latter are invested with the most desirable traits of its parents. In addition,  
progeny from a hybrid cross may possess new qualities resulting from the  
10 combination of the two parental types, such as yield enhancement resulting  
5 from the phenomenon known as heterosis. Controlled cross-fertilization to  
produce hybrid seeds has been difficult to achieve commercially due to  
competing self-fertilization, which occurs in most crop plants.

15 Hybrid seed production is typically performed by one of the following  
means: (a) mechanically removing or covering the male organs to prevent self-  
10 fertilization followed by exposing the male-disabled plants to plants with male  
organs that contain the trait(s) desired for crossing; (b) growing genetically  
20 male-sterile plants in the presence of plants with fertile male organs that contain  
the trait that is desired for crossing; or (c) treating plants with chemical  
hybridizing agents (CHA) that selectively sterilize male organs followed by  
15 exposing the male-disabled plants to plants with fertile male organs that contain  
the trait that is desired for crossing. Some disadvantages to each of these  
methods include: (a) applicability only to a few crops, such as corn, where the  
male and female organs are well separated; and it is labor intensive and costly;  
30 (b) genetically male sterile lines are cumbersome to maintain, requiring crosses  
20 with restorer lines; (c) all CHAs exhibit some degree of general phytotoxicity  
and female fertility reduction. Also CHAs often show different degrees of  
effectiveness toward different crop species, or even toward different varieties  
35 within the same species.

25 A molecular genetic approach to hybrid crop production applicable to a  
wide range of crops and involves genetic male sterility is described in EPA 89-  
344029. This system involves the introduction of a cell disruption gene that is  
expressed only in the tapetal tissue of anthers thereby destroying the developing  
pollen. The resulting genetically male sterile plants serve as the female parents  
40 in the cross to produce hybrid seed. This system could be highly effective and  
desirable. However one disadvantage is that since the male sterile parent is  
45 30 heterozygous for the sterility gene which acts as a dominant trait, only 50% of  
the plants grown from the hybrid seed are fertile, the rest retain the sterility

5 gene. This situation will result in reduced pollen shed in the production field  
which may lead to reduced seed set and yield. Addition of recombinase  
technology to this hybrid scheme allows restoration of fertility to a much higher  
percentage of plants in the production field, as well as elimination of the cell  
10 5 disruption gene. Placing the male sterility gene between recombination sites  
allows it to be deleted following introduction of a variant recombinase into the  
hybrid from the male parent.

15 Another application of the disclosed variant recombinases is in making  
seedless produce. Seedlessness is desirable in consumed produce for  
10 convenience and taste. Currently "seedless" watermelon is sold that actually  
contains some developed seed and a large number of immature seed that varies  
20 in size up to that of fully mature seed. To produce these watermelon first a  
hybrid cross is made between a tetraploid maternal parent and a diploid  
pollinator. The resulting triploid seed produces self-infertile plants that are  
15 crossed with a diploid pollinator to produce seedless fruit (Kihara, *Proc. Soc.  
Hort. Sci.*, 58: 217-230, (1951)). This production scheme suffers the following  
problems: (i) Creating a tetraploid plant, which is accomplished by a  
chromosome duplication method, is difficult. Also the number of seeds per fruit  
30 on this tetraploid plant must be low since this has a positive correlation with  
seed number in the final product (Andrus, *Production of Seedless Watermelons*,  
USDA Tech. Bull. No. 1425 (1971)); (ii) good combining ability of the diploid  
pollinator and the tetraploid plant is difficult to achieve (Henderson, *J. Amer.  
Soc. Hort. Sci.*, 102: 293-297 (1977)); (iii) the triploid seeds are much inferior  
35 to regular diploid seeds in vigor and germinability (Maynard, *Hort. Sci.*, 24:  
603-604 (1989)). These problems, together with incomplete seedlessness in the  
25 final product, make the development of seedless watermelon slow and difficult.  
This ploidy-based approach to seedlessness is possible only in those few species  
where unusual euploidy plants (tetraploid and triploid for watermelon, for  
40 example) are viable.

45 30 A molecular genetic approach to seedlessness involving the disclosed  
variant recombinases is much more efficient, resulting in a more reliably  
seedless product and does not involve changes in ploidy. Thus it is more

generally applicable to a wider range of species. A recombination site/polyA-inactivated cell disruption gene regulated by a seed-specific promoter is introduced into a plant. When this plant is crossed to a plant expressing a variant recombinase, the disruption gene is activated and expressed in the seed, thereby disrupting seed development. The certainty of endosperm failure (caused by the cell disruption gene product) leading to the abortion of the whole seed is very high. In most dicots, the endosperm supplies the nutrients needed for early embryo development. Endosperm abortion invariably leads to seed abortion (Brink and Cooper, *Bot. Rev.* 8: 423-541 (1947)).

The seed-specific promoter used can be selected from the group of promoters known to direct expression in the embryo and/or the endosperm of the developing seed, most desirably in the endosperm. Examples of seed-specific promoters include but are not limited to the promoters of seed storage proteins. The seed storage proteins are strictly regulated, being expressed almost exclusively in seeds in a highly tissue-specific and stage-specific manner (Higgins et al., *Ann. Rev. Plant Physiol.* 35: 191-221 (1984); Goldberg et al., *Cell* 56: 149-160 (1989)). Also, different seed storage proteins may be expressed at different stages of seed development and in different parts of the seed.

There are numerous examples of seed-specific expression of seed storage protein genes in transgenic dicotyledonous plants. These include genes from dicotyledonous plants for bean  $\beta$ -phaseolin (Sengupta-Goplalan et al., *Proc. Natl. Acad. Sci. USA* 82: 3320-3324 (1985) and Hoffman et al., *Plant Mol. Biol.* 11: 717-729 (1988)), bean lectin (Voelker et al., *EMBO J* 6: 3571-3577 (1987), soybean lectin (Ocamuro et al., *Proc. Natl. Acad. Sci. USA* 83: 8240-8344 (1986)), soybean kunitz trypsin inhibitor (Perez-Grau and Goldberg *Plant Cell* 1: 1095-1109 (1989)), soybean  $\beta$ -conglycinin (Beachy et al., *EMBO J* 4: 3047-3053 (1985), Barker et al., *Proc. Natl. Acad. Sci.* 85: 458-462 (1988), Chen et al., *EMBO J* 7: 297-302 (1988), Chen et al., *Dev. Genet.* 10: 112-122 (1989), Naito et al., *Plant Mol. Biol.* 11: 683-695 (1988)), pea vicillin (Higgins et al., *Plant Mol. Biol.* 11: 109-123 (1988)), pea convicillin (Newbigin et al., *Planta* 180: 461 (1990)), pea legumin (Shirsat et al., *Mol. Gen. Genetics* 215:

326 (1989)), rapeseed napin (Radke et al., *Theor. Appl. Genet.* 75: 685-694 (1988)), as well as genes from monocotyledonous plants such as for maize 15-kd zein (Hoffman et al., *EMBO J* 6: 3213-3221 (1987)), barley  $\beta$ -hordein (Marris et al., *Plant Mol. Biol.* 10: 359-366 (1988)), and wheat glutenin (Colot et al., *EMBO J* 6: 3559-3564 (1987)). Moreover, promoters of seed-specific genes operably linked to heterologous coding regions in chimeric gene constructions also maintain their temporal and spatial expression pattern in transgenic plants. Such examples include Arabidopsis thaliana 2S seed storage protein gene promoter to express enkephalin peptides in Arabidopsis and Brassica napus seeds (Vandekerckhove et al., *Bio/Technology* 7: 929-932 (1989)), bean lectin and bean  $\beta$ -phaseolin promoters to express luciferase (Riggs et al., *Plant Sci.* 63: 47-57 (1989)), and wheat glutenin promoters to express chloramphenicol acetyl transferase (Colot et al., *EMBO J.* 6: 3559-3564 (1987)). Promoters highly expressed early in endosperm development are most effective in this application. Of particular interest is the promoter from the a subunit of the soybean  $\beta$ -conglycinin gene (Walling et al., *Proc. Natl. Acad. Sci. USA* 83: 2123-2127 (1986)) which is expressed early in seed development in the endosperm and the embryo.

The cell disruption gene used can be selected from a group of genes encoding products that disrupt normal functioning of cells. There are many proteins that are toxic to cells when expressed in an unnatural situation. Examples include the genes for the restriction enzyme EcoRI (Barnes and Rine, *Proc. Natl. Acad. Sci. USA* 82: 1354-1358 (1985)), diphtheria toxin A (Yamaizumi et al., *Cell* 15: 245-250 (1987)), streptavidin (Sano and Cantor, *Proc. Natl. Acad. Sci. USA* 87: 142-146 (1990)), and barnase (Paddon and Hartley, *Gene* 53: 11-19 (1987)). Most preferred for this system is the coding region of barnase which has been shown to be highly effective in disrupting the function of plant cells (EPA 89-344029).

A highly desirable seedless system is one in which fully fertile F1 seed develops, that can then be grown into plants that produce only seedless fruit. This system is economically favorable in that for each cross pollination, a large number of seedless fruits result: the number of F1 seed from one cross X the

number of fruits produced on an F1 plant. Also incorporated in this scheme are the advantages of growing a hybrid crop, including the combining of more valuable traits and hybrid vigor. This is accomplished in the same manner as described above except that the recombination site/polyA-inactivated disruption gene is expressed from a seed maternal tissue (seed coat or nucellus)-specific promoter. For example, the seed coat is the outgrowth of the integuments, a strictly maternal tissue. Therefore the hybrid cross that brings the recombination site/polyA-inactivated disruption gene together with the recombinase gene does not involve this seed coat tissue. The seed coat of the F1 seed has either recombination sites or recombinase, depending on which is used as the female parent, and thus F1 seed develop normally. After the F1 seed gives rise to a fruit-bearing F1 plant, all vegetative cells (including seed coat cells) inherit both recombination sites and recombinase from the embryo. Thus the seed coat of the F1 plant has an activated cell disruption gene.

The seed coat is an essential tissue for seed development and viability. When the seed is fully matured, the seed coat serves as a protective layer to inner parts of the seed. During seed development, the seed coat is a vital nutrient-importing tissue for the developing embryo. The seed is nutritionally "parasitic" to the mother plant. All raw materials necessary for seed growth must be imported. In seeds of dicotyledonous plants, the vascular tissue enters the seed through the funiculus and then anastomoses in the seed coat tissue. There is no vascular tissue connection or plasmodesmata linkage between the seed coat and the embryo. Therefore, all nutrient solutes delivered into the developing seed must be unloaded inside the seed coat and then move by diffusion to the embryo. Techniques have been developed to study the nutrient composition in the seed coat (Hsu et al., *Plant Physiol.* 75: 181 (1984); Thorne & Rainbird, *Plant Physiol.* 72: 268 (1983); Patrick, *J. Plant Physiol.* 115: 297 (1984); Wolswinkel & Ammerlaan, *J. Exp. Bot.* 36: 359 (1985)), and also the detailed cellular mechanisms of solute unloading (Offler & Patrick, *Aust. J. Plant Physiol.* 11: 79 (1984); Patrick, *Physiol. Plant* 78: 298 (1990)). It is obvious that the destruction of this vital nutrient-funnelling tissue causes seed abortion.

5 The disclosed tissue-specific and site-directed DNA recombination can  
be used to obtain seedless fruit production. This method is useful for the  
production of seedless watermelon, for example. A combination of gene  
expression specific for maternally inherited seed tissue and the disclosed  
10 recombinae system can be used for the production of seedless watermelon.  
The system can be universally applied to any horticultural crop in which the  
presence of seeds is undesirable and difficult to be eliminated through  
conventional breeding methods. The system also allows the normal production  
15 of F1 seeds. The ability to maintain heterosis is an advantage of producing F2  
seedless fruits.

20 The existing production of seedless watermelon indicates that seed  
development is not essential for the watermelon fruit development. However,  
conventional production of seedless watermelon using the ploidy imbalance  
trick has never been very popular due to the difficulty of overcoming the yield  
25 and production problems. Creating and maintaining the tetraploid (4n) female  
germline, and producing the triploid (3n) seeds have made the seed cost high.  
Cross-pollination is needed for the production of triploid seeds (4n X 2n) and  
seedless fruits (3n X 2n). Also triploid seed germination is usually poor due to  
30 ploidy imbalance.

20 The present approach eliminates the dependence on polyploid germplines  
and provides an efficient system for producing seedless fruit. The products of  
double fertilization of higher plants are the embryo and endosperm. The seed  
35 coat (including the integumentary tapetum) and nucellus (the tissue  
encompassing the embryo sac) are the remaining seed tissues that are  
maternally inherited. In addition to general protection, the seed coat and  
40 nucellus also play an important role in importing nutrients into the developing  
embryo and endosperm. Seed development will be aborted if this vital nutrient-  
importing mechanism of the seed coat/nucellus is debilitated. This will be  
accomplished by using the recombinae system to activate a cell-damaging gene  
45 only in these tissues. Controlling the gene activation in a maternal tissue-  
specific manner allows production of normal F1 seed, but abortion of F2 seed.  
A seed coat or nucellus promoter is coupled to a tissue-destructive (lethal) gene

5 in order to prevent seeds from forming. The destructive gene is inactive in the  
seed parent due to the presence of a blocking transcription terminator. The  
terminator is flanked by recombination sites for subsequent excision by a  
10 recombinaise-mediated recombination event. Expression of the recombinaise is  
5 also controlled by the seed coat/nucellus-specific promoter. When plants  
carrying the separate recombinaise and recombination site constructs are  
crossed, the F1 seed will be viable because seed coat/nucellus is maternal tissue,  
15 and in that tissue recombinaise and recombination sites are not combined. When  
the F1 seed is used as planting seed, the self-pollinated or out-crossed plants  
10 will produce seedless fruits or vegetables, since in seed coat/nucellus tissues  
recombinaise and recombination sites are combined, and the lethal gene is  
20 activated.

## 2. Use of Variant Recombinases for Phage Packaging

The disclosed variant recombinases can also be used to aid in phage  
25 packaging. The cloning system described herein utilizes a headful in vitro  
packaging system to clone foreign DNA fragments as large as 95 kb which  
permits the isolation of DNA fragments that are at least twice the size of those  
that can be obtained by lambda cosmid cloning. This increased cloning  
30 capacity has the following utility:

20 (1) Genes in the 45-95 kb size range and, more particularly, in the 70-95  
kb size range can now be directly cloned and genes in the 25-45 kb size range  
can be cloned more easily.

35 (2) Chromosomal "walking" and "jumping" techniques can be speeded  
up by a factor of at least two and should be more accurate because of the  
25 reduced number of contiguous segments that need to be linked together.

40 (3) The cloning system of the invention is useful as a means for the  
delivery of DNA efficiently to bacteria which otherwise do not take up DNA  
from solution well.

45 Specifically, the headful packaging system of this invention for cloning  
30 foreign DNA fragments as large as 95 kb comprises:

(a) modifying vector DNA by inserting a stuffer fragment into a blunt  
end producing site which is proximal to a pac site;

(b) digesting the product of step (a) to produce two vector arms each of which contains (i) a blunt end, (ii) another end which is compatible with the foreign DNA fragment which is to be cloned, and (iii) a recombination site;

(c) ligating the foreign DNA to the product of step (b) without generating concatemers;

(d) reacting the product of step (c) with pac cleavage proficient extract and head-tail proficient extract wherein the ratio of large heads to small heads in the head-tail extract is at least 5:1;

(e) infecting a bacterial strain expressing a variant recombinase with the product of step (d); and

(f) recovering the cloned DNA.

The term pac is a generic name which refers to the site needed to initiate packaging of DNA. The pac cleavage proficient extract contains the recognition proteins necessary to cleave the pac site and, thus, initiate packaging. The head-tail proficient extract contains the heads and tails needed to package the cloned DNA into a virus particle. The term concatemer means a DNA molecule consisting of repeating units arranged in a head-to-tail configuration. The term stuffer fragment refers to a DNA fragment which is inserted into the vector DNA at a unique site, and within which headful packaging is terminated. The terms bacteriophage and phage are used interchangeably herein.

Although many of the elements described herein pertain to the P1 bacteriophage cloning system, those skilled in the art will appreciate that, with the exception of the components needed to package DNA (pac and packaging extracts), many of the elements discussed below, such as plasmid replicon and a multicopy or lytic replicon, pertain to the recovery of packaged DNA and can be used to recover DNA in bacteria, such as *E. coli*, with other cloning systems, for example, bacteriophage, yeast, etc.

Bacteriophages which are suitable to practice the invention must have a large head capacity and the elements necessary for packaging DNA must be defined. For example, for phages P22 and T1, which utilize headful packaging, the necessary packaging elements are defined. However, P22 and T1 do not

5 have a very large head capacity. On the other hand, for phage T4, which has a large head capacity, the necessary packaging elements have not been defined.

The elements necessary for packaging DNA (i.e., an in vitro headful packaging system) are the following:

10 5 (1) a unique site, pac, which is cleaved by recognition proteins; it is the pac cleavage proficient extract which contains the recognition proteins necessary to cleave the pac site; and

15 (2) empty phage heads into which the DNA is packaged until the head has been completely filled, then a cleavage event is triggered (the "headful" cut) which separates the packaged DNA away from the remaining components; it is the head-tail proficient extract which contains the heads and tails needed to package the cloned DNA into virus particle.

20 Although initiation of packaging is site-specific (cleavage of pac site initiates packaging), termination of packaging is not site-specific. In other words, no unique site is recognized, as packaging will terminate at whatever point the head has been filled.

25 In the case of P1, the DNA substrate used in the packaging reaction during the viral life cycle is a concatemer consisting of individual units of the P1 chromosome arranged in a head-to-tail manner. Headful packaging, using either P1 phage or any other phage, is a four step process: (1) In the first step a unique site, pac, is recognized and cleaved by the pac recognition proteins (PRPs); (2) DNA on one side of the cleavage is packaged into an empty phage head until the head has been completely filled; (3) a second cleavage event is then triggered (the "headful" cut) that separates the packaged DNA away from the rest of the concatemer; and (4) initiation of a second round of DNA packaging from the free end generated by the previous "headful" cut--hence the term processive headful cutting. However, if a concatemer is not generated then processive headful packaging does not occur.

30 The ends of the packaged P1 DNA do not contain complementary single-stranded sequences, as do the ends of packaged bacteriophage lambda DNA, and consequently after P1 DNA is injected into a bacterium its cyclization does not occur by strand annealing but rather by recombination

between homologous sequences present at the ends of the molecule. Because of this circumstance, any vector that uses P1 packaging, or for that matter any headful packaging mechanism, must devise a means of cyclizing the linear packaged DNA by recombination. Cyclizing is accomplished by incorporating recombination sites into the vector and using a disclosed variant recombinase to cyclize the DNA after injection into gram-negative bacterial strains expressing the variant recombinase.

P1 produces two head sizes, a big head that can accommodate 105-110 kb of DNA, and a small head that can accommodate no more than 45 kb of DNA. Normally the ratio of big to small heads in a P1 wild-type infection is 10:1, however, in the *cm-2* mutant of P1 used to prepare some of the packaging lysates described herein, the ratio of head sizes is 1:1. The head-tail packaging lysate prepared from the *cm* mutant of P1 contained the usual ratio of big to small heads which is about 10:1. This is the preferred lysate for preparing head-tail packaging extract. To ensure packaging of DNA exclusively into the big phage heads, the DNA must be bigger than that which can be accommodated by the small heads. It is generally desired that there be a large excess of big heads. However, the ratio of large heads to small heads should not fall below a ratio of about 5:1.

#### Illustration

The following illustration describes an example of how the disclosed method can be used to generate variant FLP recombinases with altered site specificity. As with other recombinase, the method preferably uses the following components:

1. An in vitro mutagenesis system;
2. A recombinase expression plasmid that allows varying levels of expression by a simple environmental control (for example, by the presence of varying amounts of an inducer substance in the growth media, by temperature, or by osmolarity);
3. An indicator/selector bacterial strain. The strain carries both an indicator recombination substrate for detection of recombination at the wild type recombination site and a second recombination substrate that allows

5 selection for recombinase mutants that have gained the ability to recognize and  
perform recombination at a target mutant recombination site (that is, a variant  
10 recombination site). Importantly, the wild type and target mutant sites are  
designed so that recombination between the mutant and wild type sites is  
5 blocked even with a mutant recombinase that can recognize both the wildtype  
site and also the target mutant site. This design prevents unwanted  
15 recombination between the wildtype and target mutant recombination sites that  
could interfere with either selection or detection of desired recombinational  
outcomes. The block is imposed by designing the wildtype and mutant sites to  
10 have different spacer regions (that is, different compatibility sequences), for  
example, the normal "wt" spacer for the wildtype recombination site, and an  
20 alternative spacer "A1" for the other recombination substrate. In an otherwise  
nonmutant recombination site DNA recombination proceeds efficiently both for  
recombination sites having the wt spacer (that is, a recombination between two  
25 wt sites) and also for sites having the A1 spacer (that is, a recombination  
between two A1 sites). Yet, recombination between the A1 site and the wt site  
is blocked (that is, recombination between a wt site and an A1 does not occur).  
This strategy is applicable to all recombinases that have a recombination target  
30 site displaying one or more recombinase binding sites (repeat elements) on each  
20 side of a spacer region in which recombination occurs (Nunes-Döby et al., *Nucl.*  
*Acids Res.*, 26:391-406 (1998)). Such sites display a requirement for homology  
35 in the spacer elements for optimal recombination activity and has been shown to  
be the case for members of the Int family of recombinases, including Cre,  
lambda Int, and FLP (Craig, *Ann. Rev. Genet.* 22:77-105 (1988)).

25 The preferred in vitro mutagenesis system is that of Stemmer (Stemmer,  
40 *Nature* 370:389-391 (1994)), or a variant of that strategy. After mutagenesis  
and assembly of fragments into a full-length FLP gene, it is cloned into the  
expression vector.

45 The expression plasmid to be used can be any of the "inducible"  
30 expression plasmids available in bacteria. For this illustration one of the pBAD  
plasmids for *E. coli* was chosen that allows expression of recombinase by  
growth on arabinose (Guzman et al., *J. Bacteriol.* 177:4121-4130 (1995)), and

5 which can be turned off by growth on glucose (and no arabinose). For this  
illustration the expression plasmid carries the replication origin of pACYC, and  
the FLP recombinase gene is under the control of the *E. coli* ara promoter  
10 region. In addition, the plasmid carries the selectable marker *Cm<sup>r</sup>* which  
5 confers resistance to the antibiotic chloramphenicol. Because the pACYC  
replicon is low copy, its use may be advantageous in preventing excessive  
expression of FLP. Alternatively, a higher copy replicon could be used, such as  
15 that of ColEI. In that case the expression level of FLP must also be carefully  
controlled using the inducer substance arabinose.

10 The indicator/selector bacteria carries two different reporter constructs  
for FLP-mediated recombination. The first reporter construct consists of two  
20 FRT sites (FLP recombination target; that is, the recombination site recognized  
by FLP recombinase) in direct orientation (an excision substrate) and resides on  
a low copy replicon that is compatible with the FLP expression construct. In  
25 this example the first substrate is integrated into the *E. coli* genome. This can  
be done by incorporating the FRT substrate onto phage lambda and then  
constructing a lambda lysogen. Alternatively the FRT substrate could reside on  
a low-copy replicon that is compatible with that of the FLP expression vector  
30 and which has an additional selectable marker, for example resistance to  
bleomycin. This FRT substrate carries two FRT sites in direct orientation  
20 flanking a gene whose presence can be easily monitored. In this illustration a  
constitutively expressing *lacZ* gene was used whose presence can be determined  
35 simply by growing colonies plates containing X-gal, upon which they will  
become blue in color. Loss of the *lacZ* gene by FLP-mediated recombination  
25 results in white colony formation on X-gal plates.

40 The indicator/selector bacterial strain also carries a second FRT-like  
substrate. This is a plasmid element having two FRT-M sites in direct  
orientation flanking DNA sequences (STOP) that disallow expression of a  
45 downstream selectable marker. In this example, nonexpression of the selectable  
30 marker is achieved by placing genetic elements in the following order:  
constitutive promoter – FRT-M – STOP – FRT-M – 'neo, where 'neo indicates  
the promoterless neo gene of Tn5. Hence this cassette cannot express neo and  
50

5 cells are sensitive to the antibiotic kanamycin. Excision of STOP by  
recombination at FRT-M is designed to permit expression of neo so that cells  
now become resistant to kanamycin. The plasmid carries an additional  
selectable marker, Ap<sup>r</sup> conferring resistance to ampicillin, to maintain presence  
10 of the plasmid in *E. coli*. The STOP sequence here is the strong transcriptional  
terminator rrnBT1T2 (Liebke et al., *Nucleic Acids Res.* 13:5515-5525 (1985)).

Figure 17 shows wt FRT, FRT-A1, and FRT-M sites used in this  
15 illustration. Although the wildtype FRT site displays three inverted repeat  
elements, recombination proceeds efficiently with sites carrying two of these  
repeats in the inverted configuration shown (Jayaram, *Proc. Natl. Acad. Sci.*  
*USA* 82:5875-5879 (1985)). Either the full or minimal site can be used since  
20 both are recombinationally functional. The FRT-A1 site is designed to have an  
altered spacer but which is still functional for self X self recombination  
(Senecoff et al., *J. Biol. Chem.* 261:7380-7386 (1986)). The target FRT-M site  
15 is designed to carry symmetrical mutations in the repeat elements that disallow  
efficient FLP-mediated recombination (Senecoff et al., *J. Mol. Biol.* 201:405-  
421 (1988)), and also the spacer mutation of FRT-A1.

Importantly, the FRT-M site differs from the FRT site in two ways.  
30 First, both of the 13 bp inverted repeat elements (that is the recognition  
sequences) are mutated in a symmetrical manner such that the wt FLP enzyme  
does not catalyze recombination between two FRT-M sites, or does so only  
extremely poorly (< 0.1%). Second, the spacer region (that is, the compatibility  
35 sequence; the 8 bp region between the 13 bp inverted repeats) is replaced with  
an alternate spacer A1. The alternate spacer when present in an otherwise wt  
FRT site, which we will call FRT-A1, is permissive for FLP-mediated  
25 recombination between two FRT-A1 sites, but does not permit recombination  
between FRT-A1 and FRT. Use of the FRT-M site which contains the  
heterologous spacer prevents FLP-mediated recombination between FRT and  
FRT-M by FLP or a mutant FLP protein that might otherwise catalyze  
40 recombination between the wt FRT substrate and the mutant target FRT-M  
substrate. Unwanted recombination between the wildtype and target mutant  
recombination sites would decrease efficiency of the selection procedure by (a)

5 not limiting recombination at the target mutant site specifically to these sites  
and thus compromising the selection (at FRT-M sites) for mutant FLP  
recombinases, (b) affecting the accuracy of the specific indication of activity at  
the wt FRT sites, and (c) decreasing either the plasmid stability of the FRT-M  
10 selector substrate or the integrity of bacterial chromosome (or compatible  
5 plasmid) carrying the wt FRT sites.

Procedure (Figure 18): the FLP gene is mutagenized *in vitro* and then  
15 cloned into the inducible expression vector, in this case a pBAD derivative that  
places FLP under the control of the arabinose-inducible pBAD promoter. The  
20 pool of mutagenized FLP genes is transformed into the FRT indicator/selector  
strain which is pre-induced with arabinose and/or induced with arabinose during  
DNA transformation. Bacterial colonies are then selected to be simultaneously  
resistant to chloramphenicol (to retain the FLP expression plasmid), ampicillin  
or carbenicillin (to retain the selector plasmid) and kanamycin (to select for  
25 cells in which FRT-M X FRT-M recombination has occurred) on agar plates  
containing either arabinose (for continued FLP expression) or glucose (to  
prevent prolonged FLP expression). In some instances it may be advantageous  
to limit FLP expression to better enrich for those FLP mutants that either have  
30 more avidity for recombination at the FRT-M sites or to better exclude those  
20 FLP mutants that retain activity at the wildtype FRT sites. This is because  
prolonged or high-level FRT expression can lead to inefficient but detectable  
recombination at mutant sites.

35 Either all Kan<sup>r</sup> colonies or only those that are blue on X-gal plates are  
then pooled and harvested for DNA preparation. A second round of FRT gene  
25 mutagenesis and selection is then initiated by PCR amplification. Multiple  
40 rounds of mutagenesis and selection are used to obtain FRT mutants with  
altered site-specificity. Comparison of various individual isolates allows  
determination of critical amino acid residues that contribute to the desired  
mutant phenotype.

45 30 The same rationale and procedure can be used to generate a second class  
of altered FLP recombinases. The target mutant FRT site used is, however,

different. In this case the target is the FRT-M2 site (Figure 19) which carries a different binding site mutation(s) than does the FRT-M site as described above.

#### Examples

##### Example 1: Selection of Variant Cre Recombinases

The following example describes the production and analysis of some examples of the disclosed variant recombinases. Cre mutants characterized by a wider substrate recognition were created, applying a technique called directed molecular evolution: Multiple rounds of a random mutagenesis procedure (DNA shuffling; Stemmer, W. P. C., *Proc. Natl. Acad. Sci. USA*, 91:10747-10751 (1994)) and a sensitive selection for the desired phenotypes allow to accumulate candidate mutants within the generated pools of mutated sequences. The Cre mutants created in this example showed wt-like activity on *loxP* sites. In addition, they performed on an altered substrate, called *loxK2*, that is not recognized by the wt enzyme. Two transversions from adenine (*loxP*) to thymine (*loxK2*) at positions 11' and 12' of the *lox* sequence are the barriers that inhibit wt Cre from recognizing *loxK2*: The two thymines are believed to cause repulsive forces with the acidic side chain of a glutamate residue in the J helix of wt Cre (position 262). This glutamate was found to be replaced by a glycine in all mutants with remarkably increased activity on *loxK2*. Additional site-directed mutagenesis experiments, confined to the glutamate at position 262 of Cre, could confirm that E262G but also E262W mutations alone are sufficient to increase *loxK2* activity by a factor of  $10^3$  without affecting *loxP* recognition. Other point mutations identified in the analyzed mutants may however be responsible for increasing the newly obtained specificity even further (10 fold compared to E262G alone).

#### MATERIALS AND METHODS

##### General Procedures

##### Standard Reagents

The following reagents were used in all experiments: 10x' TBE (Tris-Borate-EDTA, pH 8.3) was purchased from Biofluids, Inc. (Rockville, MD) and diluted to 1x with deionized water prior to use. TE (Tris-HCl 10 mM, EDTA 1 mM, pH 8.0 and pH 7.5) and Tris-HCl (1 M, pH 7.5 and 8.0) came from

Quality Biological, Inc. (Gaithersburg, MD), as well as autoclaved LB (Luria-Bertani) and SOC broth. L-(+)-arabinose (>99%) was ordered from Sigma-Aldrich Fine Chemicals (St. Louis, MO) and anhydrous D-glucose from Mallinckrodt Laboratory Chemicals (Phillipsburg, NJ).

#### **Gel Electrophoresis**

For DNA electrophoresis, 0.8% agarose TBE gels were used (GTG Sea Kem Agarose (FMC, Rockland, ME)). Gels were prestained with 0.25 µg/ml EtBr (Ethidium Bromide, 10 mg/ml (Life Technologies, Inc., Grand Island, NY)). The used electrophoresis apparatus was a DNA SUB CELL™ (BioRad, Hercules, CA) with an OSP 105 (OWL, Woburn, MA) powersupply. Gels were run at 60 V (5 V/cm) as recommended by Sambrook et al., Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press (Second Edition) (1989). Occasionally, for small amounts of samples, 50 ml minigels were used under similar conditions (Hoefer HE33, Hoefer Scientific Instruments, San Francisco, CA). Molecular weight standards were  $\lambda$  / Hind III digest (Research Genetics, Huntsville, AL) and Ready-Load™ 100 bp DNA ladder (Life Technologies), providing a standard size range from 100 bp to 23130 bp (Figure 4). For standard fragment purification from gel, the Geneclean II® Kit (BIO 101, Inc., La Jolla, CA (11/98) was used, following the manufacturer's instructions.

#### **Minipreps and Plasmids**

Plasmids for diagnostics, cloning, and sequence analysis were prepared using the Wizard™ Minipreps Plus Kit (Promega, Wizard™ Minipres Plus DNA Purification System. Instruction Manual (Madison, WI) (1/96)). Useful ones were assigned a pBS number and stored in TE pH 8.0 at +4°C.

#### **Oligonucleotides**

All oligonucleotides used as PCR primers, for plasmid construction, or in the mutagenesis procedure, were ordered from Midland, Inc. (Midland, TX) in gel filtration (GF) quality. The lyophilized oligonucleotides were assigned a BSB number, suspended in HPLC grad water (Sigma-Aldrich) at a final concentration of 300 µM, and stored at -20°C.

### DNA Digests and Ligations

All enzymes used for DNA manipulations (restriction enzymes, T4 ligase, etc.) were purchased from New England Biolabs, Inc., Catalog. (Beverly, MA) (1998/99) and used as recommended in the manufacturer's catalog (1998/99). Briefly, for restriction enzyme digests the total reaction volume was 20  $\mu$ l with approximately 10 units (U) of enzyme. For DNA fragment ligations, 10  $\mu$ l with 200 U of T4 DNA ligase were used.

### *E. coli* Strains

All *E. coli* strains, except otherwise mentioned, were derived originally from DH5 $\alpha$ : *endA1 hsdR17* ( $r_{KM}^{+}$ ) *supE44 thi-1 recA1 gyrA* ( $\text{Nal}^R$ ) *relA1*  $\Delta(\text{lacIZYA-argF})$  U169 *deoR* (M80 *dlac*  $\Delta(\text{lacZ})$ M15) (Woodcock et al., *Nucl. Acids Res.*, 17:3469-3478 (1989); Raleigh et al., In *Current Protocols in Molecular Biology*, eds. Ausubel, F.M. et. al. (New York: Publishing Associates and Wiley Interscience). Unit 1.4 (1989). After modification with  $\lambda$  prophages or plasmids, strains were catalogued by assigning them a BS number and stored at -80°C with 10% DMSO (Dimethylsulfoxide) after overnight culture in appropriate selection medium.

### Transformation of *E. coli*

For all plasmid transformations of *E. coli* strains, electroporation was preferred over chemical protocols. Electorcompetent cells were made and used for electroporation as described by Smith et al., *Focus*, 12:38-40 (1990). The appropriate cell porator and cuvettes were from Life Technologies. Depending on the selection procedure after electroporation, the time in SOC medium (Smith et al., *Focus*, 12:38-40 (1990) at 37°C under agitation (Lab-Line® Orbit Environ-Shaker, Lab Line Instruments, Inc., Melrose Park, IL) prior to plating on selection medium was 1 h for ampicillin (Ap) and 2 h or more h for kanamycin (Kan) selection. For induction of *cre* expression (as described below), the transformants were cultivated in SOB (Smith et al., *Focus*, 12:38-40 (1990)) supplemented with 0.2% of L-(+)-arabinose (Sigma-Aldrich) plus 20mM of  $\text{MgCl}_2$  (referred to as induction medium) for 2.5 h and 4 h before plating on the appropriate selection media (see below). Resulting colony

5 numbers were counted after overnight incubation at 37°C in a gravity  
convection incubator (Precision Scientific, Chicago, IL).

#### *E. coli* Cultures

10 LB (as mentioned above) was used as the standard medium for all *E.*  
5 *coli* cultures (liquid or solid). For selection and screening, the appropriate  
reagents at the following concentrations were added:

15 Table 1: List of reagents used for selection and screening of *E. coli*  
cultures.

	Reagent	Concentration	Stock Solution
10	Ampicillin (Ap)	100 µg/ml	50 mg/ml in H <sub>2</sub> O
	Chloramphenicol (Cm)	27 µg/ml	34 mg/ml in EtOH
20	Kanamycin (Kan)	16 µg/ml	10 mg/ml in H <sub>2</sub> O
	X-gal	0.003%	2% in DMF (w/v)
	Z-ara	0.006%	2% in DMF (w/v)

25 15 The concentration of stock solutions, stored at -20°C and their dilutions  
in liquid LB medium or LB-agar plates is given. X-gal stands for 5-bromo-4-  
chloro-3-indolyl-β-D-galactopyranoside, and Z-ara for 5-bromo-3-indolyl-α-L-  
arabinofuranoside. All reagents, except Z-ara, were purchased from Life  
30 Technologies and aliquoted in the desired stock concentration for storage. Z-ara  
20 (Berlin and Sauer, *Anal. Biochem.*, 243:171-175 (1996)) was a generous gift  
from W. Berlin.

35 Ready-to-use solid LB-agar plates (2%) plain, or supplemented with Ap  
(100 µg/ml) were purchased from Digene, Inc. (Beltsville, MD). For all other  
reagent combinations in solid medium, plates were poured according to the  
25 needs using autoclaved 2% LB agar purchased from Biofluids.

#### Polymerase Chain Reaction (PCR)

40 Standard PCR reactions were carried out in 50 µl total volume with the  
following reagents (all, except noted, from Perkin Elmer, Foster City, CA): 1x  
45 PE buffer II (without MgCl<sub>2</sub>), 2 mM MgCl<sub>2</sub>, 250 µM of each dNTP, 0.8 µM of  
30 each primer, ca. 50 ng of template DNA, qsp. H<sub>2</sub>O (HPLC grade, Sigma-  
Aldrich) to 49.5 µl. For mutagenic PCR reactions (also referred to as error-  
prone PCR), the amount of each dNTP was reduced to 20 µM and 0.25 mM of

MnCl<sub>2</sub> added. After denaturation at 95°C for 5 min, 0.5 µl of 5 U/µl PE Ampi Taq Polymerase was added at approximately the annealing temperature of the primers. After mixing, the appropriate thermal cycles were carried out (as indicated individually below), using a PTC 200 thermal cycler (DNA Engine, MJ Research, Cambridge, MA). When finished, all PCR products immediately were loaded on an agarose gel, or separated from enzyme, nucleotides and primers by applying the Wizard™ PCR Preps Kit (Promega, Wizard™ Minipres Plus DNA Purification System. Instruction Manual (Madison, WI) (1/96)). PCR products were recovered using deionized water and stored frozen at -20°C.

#### Sequence Analysis

Sequence analysis of plasmid constructions and *cre* mutants were carried out on a PE ABI PRISM™ 310 Genetic Analyzer (Perkin Elmer) according to recommendations in the manufacturer's protocol P/N 402078 Revision A (1995) for the ABI PRISM™ Dye Terminator Cycle Sequencing Kit (Perkin Elmer). Briefly, a cycle sequencing reaction contained ca. 50 ng of template DNA in miniprep quality, 4 pmol of primer, and 8 µl of the ABI Terminator Ready Reaction Mix (Perkin Elmer), in a total volume of 20 µl, and subjected to the following conditions: (96°C, 10 s; melting temperature of primer, 15 s; 60°C., 4 min) 26 times on a PTC 200 thermal cycler (MJ Research). After removal of residual primers and dye by ethanol precipitation, the DNA was resuspended in 25 µl of ABI Template Suppression Reagent (Perkin Elmer) and denatured at 95°C for 5 min before loading the ABI Genetic Analyzer. The obtained data files were examined using ABI PRISM™ Sequencing Analysis Version 3.0 Software (1996, Perkin Elmer). Gene Jockey II (1996, Biosoft, Cambridge, UK) software was used for sequence comparison, translation, and alignments.

#### Mutagenesis Procedure

##### Substrate Preparation by PCR

The *cre* gene for the following DNase I shuffling reaction was amplified by PCR using 5' forward primer BSB436 (5' AAATAATCTAGACTGAGTGTGAAATGTCC 3') and 3' reverse primer BSB376 (5' ATATATAAGCTTATCATTTACGCGTTAATGG 3'),

5 introducing an *Xba* I and *Hind* III cloning site, respectively (underlined).  
Mutagenic and non-mutagenic PCR's were carried out: (94°C, 30 s; 52°C, 30 s;  
72°C, 90 s) 45 or 30 times, respectively. The 5' primer was designed to include  
10 the endogenous Shine-Dalgarno (SD) of *cre*, whereas its three endogenous  
5 promoters were excluded (position -17, 6; positions refer to the adenine of the  
start codon of the *cre* coding sequence as position 1). Thus, after introducing  
the resulting *cre* genes into pBAD33 (see below), expression was exclusively  
15 under control of the pBAD promoter without interference or background  
expression due to endogenous promoters. Including the SD sequence of *cre*  
10 was necessary, since pBAD33 does not contain this sequence 5' of its multiple-  
cloning-site (MCS). The 3' reverse primer was designed to be homologous to  
the 3' untranslated region (UTR) of *cre* (position 1057, 1032). Mutagenic  
20 events were therefore permitted in 1020 bp of the entire 1026 bp *cre* coding  
sequence, excluding the first two codons. For the first round of the directed  
25 evolution procedure, the wt *cre* expression plasmid pBS185 (Sauer and  
Henderson, *The New Biologist*, 2:441-449 (1990)) served as template. In  
following cycles, the pool of mutated *cre* genes from the previous round was  
used. In all experiments, both, mutagenic and non-mutagenic PCR's were  
30 carried out in parallel using the appropriate template.

#### 20 Homologous Recombination *in Vitro*

##### *DNase I digest*

35 Approximately 5 µg of the *cre* PCR product (ca. 1.1 kb) were digested  
with 0.03 U of DNase I, type IV (Sigma-Aldrich) in 20 µl total volume of 50  
mM Tris-HCl pH 7.5 plus 1 mM MgCl<sub>2</sub> for 2 to 3 minutes at room temperature.  
25 After digestion, samples immediately were loaded on a 2% minigel to separate  
the generated fragments (Figure 5). Fragments of 25 bp to 300 bp were purified  
40 from the gel by DE81 (Whatman, Maidstone, GB) extraction and ethanol  
precipitation (Sambrook et al., Cold Spring Harbor, New York: Cold Spring  
Harbor Laboratory Press (Second Edition) (1989)), before suspending in 5 µl of  
45 TE pH 8.0.  
30

### Self-Printing PCR

A 60 cycle non-mutagenic PCR (as described above) was carried out without added primers, allowing the fragments to prime themselves and thereby to undergo shuffling while reassembling. Conditions for PCR were: 94°C, 90 s; (94°C, 30 s; 45°C, 30 s; 72°C, 90 s) 60 times; 72°C, 10 min.

### Reassembling of *cre*

Since the self-priming step never yielded a single size product but rather a range of fragments between 300 bp to 2000 bp (Figure 5), the self-priming PCR mixture was diluted 1/40 in a non-mutagenic PCR mix with primers BSB376 and BSB436 (see above), and subjected to an additional 20 cycles (94°C, 30 s; 52°C, 30 s; 72°C, 90 s). This additional step lead to one product of 1.1 kb size (Figure 5).

### *cre* Expression

After digesting the linkers of the generated mutant *cre* pool with *Xba* I and *Hind* III, the fragments were ligated into the identical sites of the MCS of the *cre* expression vector pBAD33 (Figure 6). Two features favored the choice of pBAD33 as the vector to express the mutant *cre* pool for the selection procedure (see below): First, its pACYC184 derived origin of replication is compatible with the ColE1 derived ones of the plasmids used in the selection procedure. Second, pBAD33 contains the promoter of the arabinose operon (pBAD), as well as expresses the regulatory protein AraC. It is therefore possible to regulate the expression of a gene cloned into the MCS and under PBAD control, from moderately high levels to nearly complete regression, by simply changing 0.2% L-(+)-arabinose in the medium to 0.2% D-glucose (Guzman et al., *J. Bacteriol.*, 177:4121-4130 (1995) and Miyada et al., *Proc. Natl. Acad. Sci. USA*, 81:4120-4124 (1984). As indicated above, the primers for *cre* PCR were designed in order to include the endogenous SD sequence but to exclude the three *cre* promoters. *cre* expression therefore will be under complete control of the pBAD promoter. This is important for the selection procedure (see below) that was intended for few Cre molecules acting on different *lox* sites. High concentrations or long term background expression of

5 *cre* could eventually defeat the selection since wt Cre also catalyzes at very low frequencies recombination events between altered *lox* sites.

#### Plasmids and *E. coli* Strains Used for Selection and Screen

##### Mutant *lox* sites

10 5 Figure 1 compares the original *loxP* site to the two mutant sites, *loxK1* and *loxK2*, used during the described experiments. The *lox* sites with 5' *Sal* I and *Xho* I compatible, and 3' *Xba* I and *Nhe* I compatible ends were received as single stranded oligonucleotides from Midland and annealed by heating the appropriate ones together at 70°C, followed by a gradual cool down.

##### Plasmids for Selection and Screening

20 Plasmid pBS561 was constructed using three fragments: (i) the 5' modified *neo* gene derived from pBS398 (Sauer et al., *Methods*, 4:143-149 (1992), (ii) the RSVneo (Gorman et al., *Science*, 221:551-553 (1983) backbone without the *neo* gene, and (iii) the oligonucleotide-derived MCS (Figure 7).

25 15 The EGFP gene derived from pEGFP-N1 (Clontech, Palo Alto, CA) was then inserted into the MCS along with 5' and 3' *lox* sites orientated in the same direction to produce plasmids pBS568 (*loxK1*<sup>2</sup>) and pBS569 (*loxK2*<sup>2</sup>). Figure 8 summarizes the procedures used to construct plasmids pBS583 and pBS584. To restore the original *neo* reading frame without the 5' extension, the *loxK*<sup>2</sup>

30 20 cassettes containing *Mlu* I/*Kpn* I-fragments from pBS568 and pBS569 were ligated into the RSV neo backbone that contains *Mlu* I and *Bgl* II sites. The *Bgl* II – *Kpn* I junction was achieved by filling the 3' recessed end of *Bgl* II with Klenow (NEB) followed by a blunt-end ligation to the *Kpn* I end. This junction also was checked by sequencing and found to be correct. A transcriptional terminator, *rrnBT1*<sub>2</sub>, derived from pBAD33 by non-mutagenic PCR with primers BSB425 (5' ATAAGCGGCCGCTGAGCTTGGCTGTT TTGGCGG 3') and BSB426 (5' GCCGTCTCGAGAGAGTTTGTAG

40 25 AAACGCAAAAAGGC 3'), was inserted into the *loxK*<sup>2</sup> cassette 3' of the EGFP gene after digest of the its *Not* I and *Xho* I linkers (underlined). With this construct, it could be predicted that a catalyzed recombination by K1<sup>+</sup> or K2<sup>+</sup> Cre mutants between the *lox* sites would result in the excision of EGFP and the transcriptional terminator, and thereby permit the transcription of the *neo* gene

5 due to the RSV promoter, located 5'. Expression of *neo* would no longer be impaired, because it is placed under control of 5' promoter elements present in RSVneo (a Kan<sup>R</sup> rendering plasmid in *E. coli*).

10 A similar *loxP*<sup>2</sup> cassette selection plasmid also was designed (pBS613, 5 Figure 9), to be used as a control. Using this plasmid, the frequency of *loxP* recombination by Cre mutants could be determined in the same manner as used to evaluate *loxK1* or *loxK2* recombination by pBS583 and pBS584.

15 Finally, a completely different set of *loxK1*<sup>2</sup>/*K2*<sup>2</sup> cassette plasmids was created. These plasmids were no longer used for selecting mutants that 10 recognize *loxK1* or *loxK2*, but rather were used to screen for these mutants in conjunction with a different bacterial background (see below). Figure 10 summarizes the construction process for pBS601 (*loxK1*<sup>2</sup>) and pBS602 (*loxK2*<sup>2</sup>): In a first step, the *neo* resistance marker of pBS581 and pBS582 (intermediates in the construction of pBS583 and pBS584) was removed by 20 deleting the *Pvu* II fragment and thus restoring the possibility to use *neo* for a different selection procedure. Following this, the *EGFP* gene between the *lox* sites was replaced by the pUC19 (Yanish-Perron, *et. al.*, 1985) derived *lac* promoter with its 3' MCS. This pUC19 fragment was obtained by non-mutagenic PCR with primers BSB448 (5' GTCAAGCTAGCTAGCAGGTTT 30 CCGACTGG 3') and BSB449 (5' ACATTGCGGCCGCGAGATCTCCTCTA GAGTCGACCTG 3'). An *Nhe* I site 5', and a *Bgl* II and a *Not* I site 3' (underlined) were introduced thereby, which made it possible to replace the *Nhe* I – *Not* I *EGFP* fragment after linker digestion. The newly generated polylinker between the two *lox* sites permitted insertion of the *Xba* I – *Bam*HI fragment of 25 pBS481 (that carries the *abfA.st* marker gene) into its *Xba* I and *Bgl* II sites. As shown recently, *E. coli* strains expressing a recombinant  $\alpha$ -L-arabinofuranosidase gene from *Streptomyces lividans* (*abfA.st*), can be detected by eye on LB plates containing 5-bromo-3-indolyl- $\alpha$ -L-arabinofuranoside (Z-ara) Berlin and Sauer, *Anal. Biochem.*, 243:171-175 (1996). This leads to the 40 formation of an indigo blue pigment, that is similar to the classical *lacZ*/5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-gal) marker system. It could therefore be expected that *E. coli* strains expressing mutant Cre

5 recombinases which allow recombination of *loxK1* or *loxK2* should lose the  
*abfA.st* gene and should form white colonies on Z-ara plates. All other clones,  
however, should be dark blue. Using *abfA.st* instead of the well established  
10 *lacZ* screen was necessary, because the *E. coli* strain (see below) used for this  
5 construct expresses  $\beta$ -galactosidase endogenously.

*E. coli Strains for Selection and Screen*

The *E. coli* strain BS583, DH5  $\Delta$ lac ( $\lambda$ D69 *loxP*<sup>2</sup>[*lacZ* LEU2]), was  
15 chosen as the bacterial background for the selection procedure for K1<sup>+</sup> or K2<sup>+</sup>  
Cre mutants by plasmids pBS583 or pBS584. Due to the *loxP*<sup>2</sup>[*lacZ*] containing  
20  $\lambda$  prophage, Cre activity on *loxP* can be evaluated simply by using X-gal plates.  
The selection strains BS1493 and BS1494 were made by introducing the  
selection plasmids pBS583 and pBS584 into BS583 (Table 2). The *loxP*2  
25 plasmid pBS613, to be used as a control, needed to be in BS583 cells as well,  
becoming the strain BS1541 (Table 2).

15 For the screening plasmids pBS601 and pBS602, *E. coli* strain NS2300  
(Sternberg et al., *J. Mol. Biol.*, 187:197-212 (1986)) was selected as host: K12  
*recA::Tn10* ( $\lambda$ i<sup>434</sup> *loxP*2[*neo*]). This strategy combines a kanamycin selection  
for Cre enzymes, that are no longer active on *loxP* (P<sup>-</sup>), with a screen for K1<sup>+</sup> or  
30 K2<sup>+</sup> enzymes. By transforming pBS601 and pBS602 into NS2300, the P<sup>-</sup>  
20 selection strains BS1523 and BS1524 were formed (Table 2).

**Selection and Screen for CRE Mutants**

*Selection for K1<sup>+</sup>/K2<sup>-</sup> and Screen for P<sup>-</sup>*

35 After ligation of the generated mutant *cre* pool into pBAD33 for 3 h,  
BS1493 or BS1494 electrocompetent cells were transformed with 2  $\mu$ l of the  
25 microdialyzed reaction mixture (VS membrane, Millipore™ Bedford, MA). To  
induce expression of the *cre* pool, the transformed cells were incubated at 37°C  
40 in induction medium for 2.5 h and/or 4 h under agitation (as mentioned before).  
Cultures were diluted 1/500 or 1/5000 and grown on LB plates with the  
45 following formulation: Ap, Cm, glucose, and X-gal for determining the  
transformation efficiency, referred to as non-selection plates. Dilutions of 1/5  
30 and occasionally 1/50 were grown on plates with addition of Kan, used to select  
for K1<sup>+</sup> or K2<sup>+</sup> mutants and called selection-plates. The formulation of the

plates served the following purposes: (i) Ap and Cm were added to assure that all clones contained both, selection and expression plasmid, (ii) X-gal to distinguish between  $P^+$  and  $P^-$  clones (Table 2). After overnight incubation at 37°C, blue and white colonies were counted and pools prepared for the next round of DNA shuffling. Alternatively, certain mutants were chosen for further analysis (see below).

*Selection for  $P^-$  and Screen for  $K1^+$  or  $K2^+$*

After 2.5 h and/or 4 h of expression of the mutant *cre* pool in the transformed BS1523 and BS1524 cells, usually dilutions of  $10^{-3}$  or  $10^{-4}$  were grown on LB agar plates supplemented with the same reagents as listed above, except, that Z-ara replaced X-gal to allow the  $K1^+/K2^+$  screen. Non-selection plates were used for determining the transformation efficiency, and Kan containing plates for the  $P^-$  selection (Table 2).

**Mutant Analysis**

**wt *cre* Expression Plasmid**

With fewer cycles (15) of non-mutagenic PCR on the *cre* expression plasmid pBS185 and after linker digestion, the *cre* pool obtained was cloned into pBAD33 and transformed into BS583 cells. After 2 h of *cre* expression, the transformed cells were grown on X-gal plates. After overnight incubation at 37°C, two white colonies (indicating *loxP* recombination) were picked for plasmid preparation and complete sequencing. No point mutation was found in either one, so that each could be used as a control plasmid for wt Cre expression. One of the two was selected for further use and named pBS606.

**Functional Testing**

In order to determine the frequency of *lox* recombination of isolated mutant Cre enzymes by the described selection procedure, it is necessary to separate the *cre* candidate expression plasmid (pBAD33) from the selection plasmid of the chosen Kan<sup>R</sup> candidate. Then, the cleaned expression vector can be used to retransform the appropriate selection and screening strain BS1494, as well as BS1493 and BS1541 to determine the candidate's capacity for *loxK2*, *loxK1* and *loxP* recombination under identical conditions. By comparing the resulting frequencies of Kan<sup>R</sup> of different Cre mutants and wt Cre, all treated

5 identically, one can determine quantitatively how well each chosen mutant candidate really performs on the altered *lox* sites.

Therefore, overnight cultures of candidates were grown in liquid LB supplemented with Cm and Kan for plasmid minipreps yielding a mixture of  
10 5 both the mutant Cre expressing plasmid pBAD33 and the newly Kan<sup>R</sup> selection plasmid. In order to eliminate the latter, minipreps were digested with the restriction enzyme *Aat* II which only cuts the selection plasmid. After  
15 transformation of BS583 cells with this digestion mixture and approximately 2 h *cre* expression, different dilutions were grown on LB agar plates supplemented with Cm and X-gal to select for pBAD33. Plates with Ap plus  
20 Cm were used to determine the background of contamination with uncut selection plasmid. The next day, clones obtained by the Cm selection were tested for Ap<sup>s</sup> and Kan<sup>s</sup> to confirm the elimination of the selection plasmid. A final overnight culture, followed by a miniprep procedure, yielded the unique  
25 plasmid for functional testing, as described above.

#### Sequencing

To obtain the DNA sequence of candidate *cre* genes in pBAD33, eight primers (BSB454 to BSB461, Table 3), four for each strand, were designed so  
30 that the entire gene could be sequenced in both directions.

#### 20 Site-Directed Mutagenesis

After identifying one essential mutation for the decrease in substrate specificity, the Stratagene QuickChange™ Site-Directed Mutagenesis Kit (Stratagene Cloning Systems, La Jolla, CA) was used to create *cre* mutants with  
35 mutations at the determined location, only. Using three different mutant primer sets (BSB 465 to BSB 470, Table 4), all steps were carried out as detailed in the manufacturer's instruction manual, except that electrocompetent BS1494 cells  
40 were used for transformation and mutant selection, replacing the provided XL1 blue cells. The made mutant candidates were subjected to functional testing and sequencing as detailed before.

45 30 In a different experiment, the DNA shuffling mutagenesis procedure was repeated on wt *cre* by adding one 5' phosphorylated strand of each set of mutant oligonucleotides (BSB465, BSB467 and BSB469) to the pool of small

fragments prior to reassembly. This allowed to incorporate them into the resulting *cre* pool. The desired mutations should consequently be introduced at much higher frequency than without the addition of oligonucleotides.

## RESULTS

### Establishment of the Selection Procedure

#### Selection Plasmids (pBS583 and pBS584)

To test the generated plasmids pBS583 and pBS584, a recombination event between their *lox* sites was mimicked by digesting pBS566 and pBS567 (intermediates of the pBS583/584 construction, containing only the 3' *lox* site) with *Sal* I and *Xho* I, followed by religation. This led to excision of the *EGFP* gene and terminators. After deletion of *EGFP* and *rmBT<sub>1</sub>T<sub>2</sub>*, the Kan<sup>R</sup> phenotype was observed as anticipated. In addition, the frequency of spontaneously occurring Kan<sup>R</sup> clones carrying the original plasmids was approximately 10<sup>-7</sup>. This background is inconsequential, since the transformation efficiency of BS583 cells was determined as 10<sup>7</sup> per µg of pBAD33.

The equivalent *loxP* control plasmid pBS613 was tested directly with the wt *cre* expression plasmid pBS606. After 2.5 h of *cre* expression, 94% of all clones were determined Kan<sup>R</sup> and about 6% showed blue color. Without *cre* expression, no Kan<sup>R</sup> and no white colonies were observed. This confirms that the control cell line BS1541 (Table 2) permits the combined P<sup>+</sup> selection and P<sup>-</sup> screen.

#### Screening Plasmids pBS601 and pBS602

Only pBS602 was tested before use by expression of wt Cre and a K2<sup>+</sup>/P<sup>+</sup> Cre mutant (see below). On non-selection medium, wt Cre expression resulted in more than 95%, whereas expression of the mutant Cre resulted in less than 3% of blue colonies. This indicates that excision of the *loxK2* flanked *abfA.st* marker by K2<sup>+</sup> Cre is possible. On selection medium, very few colonies could be found, since both types of *cre* have shown activity on *loxP* before (see below).

### Mutagenic vs. Non-Mutagenic PCR

The frequency of P<sup>+</sup> Cre mutants obtained after non-mutagenic and after error-prone PCR was determined by the following experiment: After one mutagenic or one non-mutagenic PCR on the wt Cre expression plasmid pBS185, the resulting *cre* pools were inserted into the expression vector pBAD33 and transformed into BS583 cells. After 2.5 h of arabinose-mediated induction or glucose-mediated repression (by SOC medium) or *cre* expression, dilutions were transferred to LB plates with Ap, Cm, glucose and X-gal. The results are presented in Table 5: Under glucose repression, exclusively blue colonies could be identified (first line in Table 5), indicating that *cre* expression is insufficient for *loxP* recombination and excision of *lacZ* of BS583. Induction with L-(+)-arabinose, however, lead to the formation of white colonies at the presented frequencies (second line). indicating that (i) the described control of *cre* expression by pBAD33 is functioning, and (ii) the mutagenic PCR conditions cause three times more impaired Cre enzymes for *loxP* recombination than the non-mutagenic conditions (60% blue colonies vs. 30%). It is worth mentioning that ligation reactions lacking *cre* insertion resulted in 50 to 100 times less blue colonies than obtained with the ligations with *cre* insertion. This phenotypically blue background of empty pBAD33 was subtracted before calculating the presented data.

Leung et al., *Technique*, 1:11-15 (1989) reported that the frequency of point mutations created by error-prone PCR is about 0.5%. If this is true, in average five point mutations should occur in each 1 kb *cre* coding sequence subjected to an error-prone PCR. By extrapolating this data to the three times less P<sup>+</sup> enzymes after a non-mutagenic PCR, one can conclude that the frequency for point mutations should also be reduced by a factor of three to 0.18%. Experiments made by Zhou et. al. (1991) showed 11% of a 633 bp marker gene phenotypically impaired after non-mutagenic PCR. About 37% of all genes in the pool, however, carried at least one point mutation. Even though the conditions for the non-mutagenic PCR were similar, the observed discrepancy between 11% and 20% of phenotypical mutants may be due to a variety of reasons, among which: (i) the size difference between the two genes

(633 bp vs. 1020 bp), (ii) different elongation times during PCR, and (iii) different sensibility of the two proteins for disabling point mutations.

#### Testing wt Cre on *loxK1*<sup>2</sup> vs. *loxK2*<sup>2</sup> Substrates

The level of *loxK1* and *loxK2* recombination due to wt *cre* expression was determined using the wt Cre control plasmid pBS606. After transformation of the cell lines BS1493 and BS1494 with pBS606 and 2.5 h and 4 h of *cre* expression, cells were grown on selection and non-selection plates (as described previously). The recombination frequency between the altered *lox*<sup>2</sup> sites was considered equal to the observed frequency of Kan<sup>R</sup> phenotype: for *loxK1*, it was about 10<sup>-5</sup> after 2.5 h and 2 X 10<sup>-5</sup> after 4 h of wt *cre* expression, for *loxK2*, it changed from about 2 X 10<sup>-5</sup> to 2 X 10<sup>-3</sup>. All colonies found were white, indicating effective *loxP* recombination by wt Cre within the allowed expression time. This result shows that long term expression of the wt enzyme permits a slight increase in recombination between the altered *lox* sites. The use of pBAD33 to avoid background *lox* recombination by suppressing *cre* expression is therefore justified. Because *loxK2* was eventually 100 fold better recognized by wt Cre than *loxK1* (but still at low frequency), first the creation of novel Cre recombinases with *loxK2* specificity was attempted.

#### Mutagenesis on *loxK2*

##### First Rounds of Directed Evolution

The result of the first four rounds of the described mutagenesis procedure on *loxK2* with selection plasmid pBS584 are presented in Table 6. The following symbols are used to describe the status of the DNA shuffling procedure for *cre*: "o" indicates a non-mutagenic PCR, "m" a mutagenic PCR, and "x" stands for the *in vitro* reassortment event. For example, mxoxox *cre* represents a *cre* pool subjected to three rounds of the directed evolution process, with a mutagenic PCR followed by *in vitro* shuffling in the first, and non-mutagenic PCRs and shuffling as mutagenic and recombinogenic events in the following two rounds. The phenotypically blue background due to empty pBAD33 was subtracted from all results by control ligations without *cre* insertion. In every round, error-prone and non-mutagenic PCR served as the necessary mutagenic event on the template pool of the previous round. After *in*

5 *vitro* reassortment and selection, always the larger Kan<sup>R</sup> population of the two  
parallel experiments was chosen as template for the next round (as indicated in  
the last column of Table 6). Only in the first round the error-prone PCR could  
lead to more candidates, whereas in all following rounds the reduced mutagenic  
10 frequency of the non-mutagenic PCR turned out to be more beneficial. The  
density of point mutations resulting out of two mutagenic PCR's was obviously  
too high to allow efficient elimination of deleterious mutations from  
advantageous ones during the *in vitro* shuffling step. This is confirmed by the  
15 high frequencies of blue colonies found within any pool in any round subjected  
to mutagenic PCR's twice. Error-prone PCR in the context of the applied  
selection therefore appears to be useful in the first round only, where its three  
times higher mutagenic frequency increases the amount of beneficial mutations  
20 compared to a non-mutagenic PCR. With increasing cycle numbers non-  
mutagenic PCR's should be preferred to avoid high densities of deleterious  
25 mutations.

The established directed molecular evolution process allows effective  
evolution of *cre*. Column five of Table 6 shows that with every round the  
number of Kan<sup>R</sup> colonies is increasing, while the time for *cre* expression could  
30 be lowered from 4 h to 2.5 h (column three). After only three rounds, Cre  
20 mutants capable for *loxK2* recombination at decreased concentrations due to the  
reduced expression time were found. However, it was not possible to isolate  
any blue colony on the selection plates. All identified K2<sup>+</sup> mutants therefore are  
35 also P<sup>+</sup>. As mentioned above, high densities of deleterious mutations in the *cre*  
pool subjected to error-prone PCR's twice could explain why no blue colonies  
25 were seen on selection plates, even with over 90% of P<sup>+</sup> candidates on non-  
selection plates.

#### Evaluation of Six K2<sup>+</sup>/P<sup>+</sup> *cre* Mutants

##### Functional Test

45 As indicated in Table 6, 36 white Kan<sup>R</sup> colonies could be isolated from  
30 the mxoxox *cre* pool after only 2.5 h of *cre* expression with an 1/5 dilution  
grown on selection plates. This result indicated that competent Cre mutant  
capable of *loxK2* recombination were produced. Six were selected for further  
50

5 analysis. After elimination of the selection plasmid from the minipreps (as described in Materials and Methods), all six of them, as well as wt Cre (pBS606) were subjected to the described functional test on *loxK2*, but also on *loxK1* and *loxP* recombination with plasmids pBS584, pBS583 and pBS613.

10 5 The results are presented in Table 7.

Briefly, by selecting for *loxK2* recombination, all mutants except mxoxox 4, showed significantly increased percentages of Kan<sup>R</sup> (between 3% and nearly 70%), compared to wt Cre (0.002%), as indicated in column three of Table 7. This indicates a 10<sup>3</sup> to over 10<sup>4</sup> fold increase in activity on *loxK2*.

15 10 On *loxP*, all (including mxoxox 4) showed recombination frequencies between 80% and 100% after 2.5 h of *cre* expression (column 4). This was expected from the results obtained with the X-gal screen for *loxP* recombination during the selection procedure. 2.5 h of induction for wt *cre* expression is therefore sufficient for almost complete *loxP* recombination, justifying 2.5 h of  
20 15 expression of mutant *cre* pools for selecting competent K2<sup>+</sup> Cre mutants. The observed slight decrease in *loxP* recombination with the mutants mxoxox 3 to 6 either derived from usual variations during experiments, or may indicate a slightly reduced *loxP* activity. With BS1541 blue colonies on both selection and non-selection plates were found in approximately the same frequency (2% to 20%) as kanamycin sensitivity (Kan<sup>S</sup>). This indicates competition between the *loxP*<sup>2</sup>[*lacZ*] site on the genome and the *loxP*<sup>2</sup>[*EGFP-rrnBT<sub>1</sub>T<sub>2</sub>*] sites on pBS613. Since the same Cre mutants never resulted in blue colonies during selection for *loxK2* recombination in cell line BS1494, it is possible to conclude that *loxP* is still preferred over *loxK2*. This argument is supported by higher  
25 40 frequencies of *loxP* recombination, close to 100%, compared to the *loxK2* recombination frequencies of 3% to 70%.

All frequencies found for *loxK1* recombination, determined by using cell line BS1493, lie below 0.01% after 2.5 h, as well as after 4 h of induction of *cre* expression (column five). This indicates that no analyzed mutant developed an  
45 30 increased activity on *loxK1* compared to wt Cre.

5 To summarize, five out of the six analyzed mutants showed a significant decrease in specificity, resulting in the possibility for *loxP* and *loxK2* recombination.

#### Sequencing

10 5 The six described mutants have been completely sequenced in both directions to determine the mutations which lead to the observed decrease in specificity. The resulting aligned *cre* coding and aa sequences of all mutants and wt Cre are represented in Tables 10 and 11. Each mutant showed between 15 3 and 8 point mutations, altogether 31, as listed in column two of Table 8. The overall mutagenic frequency can therefore be calculated at 0.5% (31 mutations in 6 clones of 1020 bp), which is similar to only one round of error-prone PCR 20 (Leung et al., *Technique*, 1:11-15 (1989)). The reason for the low frequency of point mutations after three rounds of the mutagenesis procedure (i.e. nine PCR's) is the applied selection after each round, *cre* mutants with low density of 25 point mutations seem to be favored by the stringent kanamycin selection.

26 of the 31 identified point mutations resulted in aa changes compared to the wt sequence, as indicated in column three. No deletions or frame-shift mutations, as well as no codons affected by more than one point mutation at the same time could be identified. All possible transition events could be observed, 30 but only half of all possible transversion events (shown in Table 12). Adenine to guanine and vice versa transition events represented almost 30% of all identified point mutations. All other events occurred less frequent (10%, 7%, or never). The represented statistic may however be biased, either since only six mutants were analyzed, or due to the directed molecular evolution technique 35 itself: The types of point mutations observed less often, may more frequently be deleterious in *cre* and were consequently removed from the pools. With more mutants to be sequenced, this question could be addressed further.

40 Only one point mutation, a transition event from adenine to guanine at position 785 in the *cre* coding sequence is common for all 5 mutants with 45 remarkably increased *loxK2* activity. This mutation leads to a replacement of a glutamate residue at position 262 in the J helix of wt Cre by a glycine (indicated in column five of Table 8). This glutamate is believed to contact the *loxP* 50

sequence at positions 11 or 12 with its acidic side chain (Guo et al., *Nature*, 389:40-46 (1997)). Another point mutation, resulting in a conservative threonine to serine exchange at position 316 was identified in three mutants. Five point mutations were found independently in two of the six mutants, among which two types of silent mutations. Finally, eleven point mutations occurred only once (indicated in column four). Therefore, the critical mutation for *loxK2* activity appears to be E262G. Some of the additional mutations could be responsible for the observed ten fold difference in *loxK2* recombination frequency among the five E262G carrying mutants.

#### Site-Directed Mutagenesis to Verify Results

##### Site-Directed Mutagenesis Procedures

To determine whether the E262G mutation alone is responsible for the increase of *loxK2* activity of the Cre mutants by at least a factor of  $10^3$ , two different experiments were made:

First, the described directed evolution procedure was repeated in three different sets on wt *cre*, by adding three 5' phosphorylated mutant oligonucleotides prior to reassembly: Incorporation of the first oligonucleotide (BSB465) into *cre* should lead to the E262G mutation, incorporation of the second one (BSB467) to a E262A mutation, and the equimolar mixture of random oligonucleotides (represented by BSB469) to  $20^3$  (= 8000) possible aa combinations at positions 261 to 263 of Cre. According to Stemmer, W. P. C., *Proc. Natl. Acad. Sci. USA*, 91:10747-10751 (1994), these oligonucleotides should be incorporated during reassembly and cause the desired mutations at a frequency of about 8%. After insertion of the resulting *cre* pools into pBAD33 and 2.5 h of expression, 0.8% of white Kan<sup>R</sup> colonies were found with BSB465. Therefore, the frequency of the Kan<sup>R</sup> phenotype due to *loxK2* recombination is ten times lower than the expected frequency for the E262G mutation to occur. This indicates that the E262G mutation increases the specificity for *loxK2* from 0.002% of wt Cre (see above) to approximately 10% of recombination during the standard expression experiment (2.5 h of *cre* expression prior to selection). By using the second oligonucleotide (BSB467) only 0.09% of the total white colonies showed Kan<sup>R</sup>. Thus, E262A still favors *loxK2* recombination but by

5 almost a factor ten less efficiently than the E262G mutation. With the random  
oligonucleotide mixture (BSB469), the frequency of *loxK2* recombination  
shrank to 0.02%. Compared to a control experiment with no oligonucleotides  
10 5 added, resulting in 0.003% Kan<sup>R</sup> (consistent with the results obtained below),  
still some of the possible 8000 aa combinations at positions 261 to 263 of Cre  
are expected to favor *loxK2* recombination. Since the frequency of blue  
colonies on non-selection plates was about 30%, a clear indication for additional  
15 mutations, it cannot be completely excluded that some of the occurred Kan<sup>R</sup>  
mutants of this experiment have been carrying some additional beneficial  
10 mutations. Nevertheless, this experiment indicated that the E262G mutation is  
probably the basis for significant increase in *loxK2* activity.

20 For better defined mutations, a second experiment was done: The same  
oligonucleotides, now in sets for both DNA strands, were used with the  
Stratagene QuickChange™ Site-Directed Mutagenesis Kit as described in  
15 Materials and Methods. After transformation of the *loxK2* selection strain  
BS1494, the percentage of white Kan<sup>R</sup> colonies in the three different  
experiments could be determined as 6.2% for E262G, 0.8% for E262A and  
0.6% for the 8000 different ten (10) days combinations as positions 261 to 263.  
30 Due to the QuickChange™ procedure, which eliminates parental DNA, one  
20 could expect that the desired mutations occurred in at least 50% of all clones of  
the three different pools. This assumption indicates that the E262G mutation  
alone results in approximately 12% or less recombination frequency on *loxK2*  
35 (6% out of 50% or more carrying the mutation) and the E262A exchange in  
1.6% or less (under standard conditions). This calculation is not valid for the  
25 third set of oligonucleotides, because no defined mutation is introduced but  
rather a mixture of 8000 different ones. Blue colonies were not found during  
40 this experiment, confirming that the frequency of additional mutations altering  
Cre activity on *loxP* was very low.

45 The results of both experiments indicate consistently, that the E262G  
30 mutation alone is sufficient to increase *loxK2* recognition by approximately a  
factor of 10<sup>3</sup> compared to wt Cre. The ten times higher frequency observed  
with three of the six analyzed mutants after three rounds of the mutagenesis

5 procedure can only be explained with additional beneficial mutations. The E262A mutation increases the frequency of *loxK2* recombination also, but approximately by a factor of ten less effectively than E262G.

#### Sequencing and Elimination of Possible Additional Mutations

10 5 Two white Kan<sup>R</sup> colonies of each pool derived from the site-directed mutagenesis procedure were selected and the entire *cre* gene was sequenced. Both E262G and E262A candidates showed the desired sequence with no additional mutations. The random candidates surprised: One of them did not reveal any point mutation, representing an artifact which managed to survive the selection, whereas the other one showed nucleotide alterations from position 783 to 786. The wt sequence (<sub>781</sub>CTG GAA<sub>786</sub>) was found to be changed to (<sub>783</sub>CTT TGG<sub>786</sub>), resulting in one silent mutation, conserving L261, and a E262W exchange. To exclude possible mutations in pBAD33 due to the PCR involved site-directed mutagenesis procedure, the three defined *cre* mutants (E262G, 15 E262A, and E262W) were excised by *Hind* III and *Xba* I and reinserted into the MCA of fresh pBAD33.

#### Functional Testing

30 The three defined Cre mutants for the amino acid position 262 were subjected to a functional test for *loxK2*, *loxK1* and *loxP* recombination activity, as mentioned before. The results are summarized in Table 9. First, the results described previously for the E262G and E262A mutations were confirmed: As indicated in column three, the *loxK2* recombination frequency increased 10<sup>3</sup> fold with the E262G mutant compared to the wt enzyme, whereas the E262A mutant shows only an increase of 200 fold. Surprisingly, the E262W mutant 25 also achieved a similar activity on *loxK2* as seen with the E262G mutant. The test for *loxP* recombination frequency with cell line BS1541 (column four) showed that the ability for *loxP* recognition is at best slightly impaired by the three different mutations (as already seen with the analyzed mxoxox mutants). Again, blue colonies could only be found with this cell line, indicating that *loxP* 45 is preferred over *loxK2* (as described before). As expected, none of the three mutants performed significantly better on *loxK1* than wt Cre (column five). To conclude, this final experiment provides the necessary evidence that the E262G

5 mutation presents the basis for the observed decrease in specificity of Cre.  
However, additional mutations seem to be helpful to increase the newly  
obtained activity further. In addition, glycine at position 262 is not the only  
possible residue to permit a remarkable increase in *loxK2* activity.

## 10 5 DISCUSSION

### K2<sup>+</sup>/P<sup>+</sup> Cre Mutants

The chosen random mutagenesis procedure linked to the described  
15 selection in *E. coli* allowed the identification of Cre mutants, characterized by a  
wider substrate recognition. The evolved enzymes showed wt-like activity on  
10 *loxP* sites and in addition had almost the same activity on altered *lox*<sup>2</sup> sites,  
referred to as *loxK2*. By contrast, the wt enzyme showed only marginal activity  
20 on *loxK2*<sup>2</sup> substrates.

### *lox* Sites

*loxP* and *loxK2* differ at several locations as illustrated in Figure 1:

15 First, the three outermost bp of the inverted repeats are altered, facilitating the  
construction of the various plasmids used for selection. Second, the entire non-  
canonical 8 bp spacer is completely exchanged, and third, two transversion  
events (thymine to adenine) are introduced at positions 11 and 12 of the *lox* site.  
20 Only the two transversion events are considered important for inhibiting wt Cre  
from recognizing the site. Other investigations have shown before, that the two  
alterations of *loxP* mentioned first are without inhibiting effect on wt Cre (B.  
Sauer, unpublished results). The design of *loxK2* was based in part on so-called  
35 cryptic *lox* sites, that were identified previously in the yeast genome (Sauer, B.,  
*J. Mol. Biol.*, 233:911-928 (1992)). Another consideration was the choice of a  
25 good starting sequence for the described mutagenesis procedure. Starting with  
sites that contained several and/or widespread alterations of *loxP* was avoided,  
because the greater the number of alterations in the substrate, the more the  
enzyme would have to be altered. Consequently, to most effectively use the  
mutagenesis procedure, the two sites presented in Figure 1, *loxK1* and *loxK2*,  
40 were designed to have only two critical alterations. In initial experiments, wt  
Cre was found to recombine *loxK2*<sup>2</sup> substrate pBS584 slightly better than  
30 *loxK1*<sup>2</sup> pBS583. This difference may depend on the fact that in *loxK2* the two

alternations are located next to each other, while in *loxK1* they are separated by 3 bp (positions 14 and 10). Thus, *loxK1* could interfere with wt Cre binding at two distinct DNA-protein interaction sites as compared to *loxK2*, where only one location of incompatibility is available. For this reason, *loxK2* was chosen for the initial set of experiments, described in this work.

#### Cre Mutants after Three Rounds of Directed Evolution

Three iterations of the *in vitro* evolution procedure were necessary to identify 36 candidates, expressing Cre mutants that could process *loxK2* (based on the applied selection procedure in *E. coli*). Tests showed that five out of six selected ones had  $10^3$  and  $10^4$  fold increased activity on *loxK2* when compared to wt Cre. On *loxP* and *loxK1*, however, there was almost no difference between wt and the mutant enzymes. The mutants therefore had developed an increased tolerance from transversions at positions 11 and 12 (*loxK2*) of the *lox* sequence, but not for other positions like 10 and 14 (*loxK1*). To indicate this phenotype they were referred to as  $K2^+/P^+$ .

#### The E262G Mutation

Sequence analysis identified that the five mutants with remarkably enhanced *loxK2* activity ( $10^3$  to  $10^4$  fold compared to wt Cre) had in common only one point mutation, leading to the aa change E262G. Site-directed mutagenesis experiments confirmed, that the E262G mutation is sufficient to increase *loxK2* activity by a factor of  $10^3$  compared to wt Cre.

Based on the recently described crystal structure (Guo et al., *Nature*, 389:40-46 (1997)) glutamate at position 262, located in the J helix of the enzyme, may be a DNA contacting residue and permit the formation of a hydrogen bond between the carboxyl group of its side chain and an amino group of one of the two adenines at positions 11' or 12' in the *loxP* sequence (Figure 1). However, changing these two bases to thymines in the *loxK2* sequence should lead to an electrostatic repulsion between their oxygens and the acidic side chain of glutamate. This could explain the observation that wt Cre is unable to catalyze a recombination between two *loxK2* sites. Exchanging glutamate for a glycine, that does not have a side chain, should remove the electrostatic repulsion and thereby permit *loxK2* binding. On the other hand,

5 this alteration could affect *loxP* binding because an electropositive DNA-protein interaction may be lost. Results from the mentioned experiments support this proposition: The E262G mutation alone lead to an increase in *loxK2* recombination from 0.002% to about 2% and to a slight decrease in *loxP* recombination from 94% to about 80%. The 50 times lower frequency of *loxK2* compared to *loxP* recombination may depend on the second thymine or more likely on the complementary adenine (position 11 in *loxK2*) that could contribute to sterical repulsion between the J helix and the *loxK2* site. To prove this hypothesis, the role of arginine at position 258, located one helix turn away from the glutamate, should be further investigated by site-directed mutagenesis. As proposed by Guo et al., *Nature*, 389:40-46 (1997), R258 is a DNA contacting residue that forms hydrogen bonds with the guanine at position 10' of *loxP* and may also interact with the bp at position 11. There is yet no confirmatory experimental evidence for this proposal.

15 Results from three initially isolated mutants (mxoxox 1 to 3) indicated about 50% of recombination frequency on *loxK2*. This is about ten fold higher than that obtained with the E262G mutation alone. It is therefore likely that some of the additional point mutations identified in these three mutants account for this increase in activity. Table 8 lists all point mutations that were found. If silent and conservative mutations are considered not to influence specificity, only a limited number of candidates to account for the phenotype remains. Among these, S254G and Q255R of mxoxox 2 and 3, because of their location close to the amino-terminus of the J helix, could be expected to influence DNA contacts with positions 11 or 12 of the *lox* site. The other mutations are scattered in the N- and C-terminal domain of Cre. All, except R101Q of mxoxox 5, affect aa that are not located within the proximity of DNA contacting areas. Some appear independently in two mutants, e.g. D29A or D189N, that could influence protein folding or the interactions among the four Cre enzymes necessary for recombination. Such alterations could influence for example the orientation of the J helix and thereby reduce remaining interference between the *loxK2* site and the enzyme. Alternatively, some mutations, also silent ones, could influence protein expression, leading to a faster accumulation of enzymes

5 and consequently to higher recombination frequency. This possibility should  
however also influence *loxP* recombination. In fact, after 2.5 h of *cre*  
expression, the mutants *mxoxox* 1 and 2 showed a slightly higher frequency of  
10 *loxP* recombination compared to wt *cre* (98% vs. 94%). This difference, on the  
5 other hand, may be attributed to the variations that occur normally within  
experiments. To address this question further, shorter *cre* expression times on  
*loxP* would be required.

15 When the coordinates of the crystal structure (Guo et al., *Nature*,  
389:40-46 (1997)) are available (protein data bank, Brookhaven National  
10 Laboratory), it will be possible to confirm many of the tenets discussed.

20 Finally, the increase in *loxK2* tolerance between the E262G mutation  
alone and the isolated mutants carrying additional point mutations justifies the  
use of the DNA shuffling procedure linked to selection: Not only has it  
permitted the elimination of deleterious mutations from the sequence pool, but it  
15 helped to accumulate various more or less beneficial aa alterations as well.

#### The E262A and E262W Mutations

The mentioned site-directed mutagenesis procedure was used to generate  
two more defined Cre mutants, E262A and E262W. Compared to the E262G  
30 mutation, E262A permitted *loxK2* recombination ten fold less effectively. The  
20 E262W mutation however resulted in similar activity on *loxK2*.

The aliphatic side chain (a methyl group) of A262 could be the reason  
35 for slight sterical interference. This would explain the observed reduced  
frequency of *loxK2* activity with E262A. *loxP* recognition, however, could not  
be found to be affected compared to E262G. The lowered *loxK2* activity  
25 explains why no E262A mutation was identified in the small pool of six  
analyzed mutants: With a ten fold decrease in activity, one would expect to  
40 encounter the corresponding mutation ten times less often during selections as  
well.

45 In contrast to the small side chain of alanine, the large aromatic side  
30 chain of W262, was expected to inhibit recombination between any *lox* sites due  
to sterical interference. Surprisingly, this seems not to be true. A possible  
explanation for the observed activity on *loxP* and *loxK2* could be that the

aromatic and planar structure of the tryptophan side chain, fits better into the J helix – *lox* interface than does a methyl group. Different influences of A262 and W262 on folding of the J helix could also contribute to the observed phenotypes. The reason why the E262W mutation was not identified among the six analyzed mutants is the genetic code. Whereas for the E262G and E262A mutations only one bp in the glutamate encoding GAA codon needs to be mutated to GGA or GCA, for E262W the whole codon must be exchanged to TGG. This is unlikely to occur during the applied random mutagenesis procedure with a mutation frequency of 0.5%. Other amino acid changes due to two or three mutations of the glutamate encoding codon therefore cannot be considered to have occurred during the random mutagenesis procedure.

When the coordinates of the crystal structure are public, it will be interesting to confirm and further investigate the discussed hypothesis.

#### ABBREVIATIONS

5-BI-ara a.k.a Z-ara	5-bromo-3-indoyl- $\alpha$ -L-arabinofuranoside
aa	amino acid
Ap(Ap <sup>R</sup> , Ap <sup>S</sup> )	ampicillin
bp	base pairs
Cm(Cm <sup>R</sup> , Cm <sup>S</sup> )	chloramphenicol (resistant, sensitive)
<i>cre</i> – Cre	cyclization recombination ( <i>gene</i> – Protein)
Da	Dalton
DMSO	dimethylsulfoxide
ds	double stranded
GF	gel filtration
K1 <sup>+</sup> or K2 <sup>+</sup>	Cre mutant capable for <i>loxK1</i> or <i>loxK2</i> recombination
Kan (Kan <sup>R</sup> , Kan <sup>S</sup> )	kanamycin (resistant, sensitive)
<i>lox</i>	locus of crossover
MCS	multiple cloning site
mRNA	messenger RNA
<i>neo</i> – Neo	neophosphotransferase ( <i>gene</i> – Protein)
P <sup>-</sup> or P <sup>+</sup>	Cre mutant defective or capable for <i>loxP</i> recombination

5	PCR	Polymerase Chain Reaction
	SD	Shine-Dalgarno sequence
	ss	single stranded
	TK	thymidine kinase
10	5 wt	wildtype
	X-gal	5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside

### 15 Example 2: Analysis of Variant Cre Recombinases

This example describes analysis the activity of several specific variant  
10 Cre recombinases.

#### Vectors

20 *pBS606, 614, 626, 627, 628 and 650*: pBAD33 with wt, E262G, E262G/D29A, E262G/D189N, E262G/T316S, and R3M3 *cre* insertion used for expression of the corresponding Cre proteins in DH5 $\alpha$  for *in vivo* testing.

25 15 *pBS632 to pBS641*: pUC19 based plasmids for *in vivo* tests of different Cre mutants on a variety of different *lox* sites and combination of *lox* sites, all bearing the FAS1 spacer (Sauer B., *Nucleic Acids Res.*, 24:4608-4613 (1996)). Recombination between two *lox* sites leads to excision of a *neo* cassette to give  
30 kanamycin sensitive *E. coli*. The same plasmids were also used for *in vitro* recombination experiments.

35 *pRH200*: wt Cre expression plasmid (a generous gift from Ron Hoess, DuPont, Wilmington, DE) used to overexpress wt Cre in BL21(DE3) (Novagen, Madison, WI) strain.

40 25 *pBS654 to pBS658*: wt *cre* of pRH200 was replaced with different *cre* mutants (E262G, E262G/D29A, E262G/D189N, E262G/T316S, and R3M3) using *Age* I and *Mlu* I restriction sites.

#### *E. coli* Strains

45 30 *BS583*: The *E. coli* strain BS583, DH5  $\Delta$ lac ( $\lambda$ D69 *loxP*<sup>2</sup>[*lacZ* LEU2]), was chosen as the bacterial background for the selection procedure using plasmid pBS584. Due to the *loxP*<sup>2</sup>[*lacZ*] containing  $\lambda$  prophage, Cre activity on  
50 *loxP* can be evaluated simply by using X-gal plates.

5                    *BS1494*: The *E. coli* strain for selection was established by introducing  
the selection plasmid pBS584 into BS583. Thus, BS1494 allows a kanamycin-  
selection for *loxK2* and in parallel a blue/white-screen for *loxP* recombination  
10                    with 5-bromo-4chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-gal). Note that the  
5                    spacer region of the *loxK2* site (FAS1) is different from the original one of *loxP*.  
Thus, recombination events between *loxP* of the  $\lambda$  prophage and *loxK2* of the  
selection plasmid pBS584 catalyzed by potent Cre mutants are excluded in  
15                    BS1494.

*BS1576 to BS1581*: For the *in vivo* recombination experiments wt and  
10                    mutant Cre expressing strains were generated by introducing plasmids pBS606,  
614, 626, 627, 628, and 650 into DH5 $\alpha$ .

#### 20                    Transformation of *E. coli*

                    For plasmid transformations of *E. coli* strains, electroporation was  
preferred over chemical protocols. Electrocompetent cells were made and used  
25                    15                    for electroporation as described by Smith et al., *Focus*, 12:38-40 (1990). The  
appropriate cell porator and cuvettes were from Life Technologies (Bethesda,  
MD).

#### 30                    Site-Directed Mutagenesis

                    The QuickChange<sup>TM</sup> Site-Directed Mutagenesis Kit (Stratagene Cloning  
20                    Systems, La Jolla, CA) (1997) was used to generate defined single and double  
mutations in the *cre* gene.

#### 35                    Overexpression and Purification of Candidate Mutants

                    Wild-type and five different mutant Cre proteins were overexpressed  
using plasmids pRH200, and pBS654 to pBS658 in conjunction with Novagen  
25                    BI21 (DE3) cells. After induction for 2.5 h cells were harvested, sonicated, and  
40                    Cre partially purified after DNase I digest with a single step Whatman® P11  
resin (Whatman Inc., Fairfield, NJ) as described before by Wierzbicki et al., *J.*  
*Mol. Biol.*, 195:785-794 (1987). The obtained Cre preps were about 80% pure  
45                    and protein concentrations ranged between 100 and 200 ng/ $\mu$ l.

#### 30                    Mutant Analysis

*In Vivo*: Plasmids pBS632 to pBS641 were transformed into Cre-  
expressing *E. coli* strains BS1576 (wt), BS1577 (E262G), BS1578

(E262G/D29A), BS1579 (E262G/D189N), BS1580 (E262G/T316S), and BS1581 (R3M3). After 1 hour of induction of *cre* expression with 0.2% L-(+)-arabinose,  $10^{-5}$  dilutions were plated on non-selection medium containing 0.2% D-glucose. After overnight incubation at 37°C, colonies were transformed to kanamycin-selection plates for the described negative selection for *neo* excision.

*In Vitro*: Purified Cre mutants were used for both *in vitro* recombination and gel retardation experiments as described before by Sauer B., *Nucleic Acids Res.*, 24:4608-4613 (1996) and Wierzbicki et al., *J. Mol. Biol.*, 195:785-794 (1987), respectively. For the recombination reactions plasmids pBS632 to pBS641 served as substrates, whereas for the DNA binding reactions  $\gamma[^{32}\text{P}]$ -dATP (Amersham Pharmacia Biotech, Piscataway, NJ) end-labeled 35 bp oligonucleotides were used, each encoding a *lox*-halfsite and one half of the FAS1 spacer.

#### Evaluation of Six K2<sup>+</sup>/P<sup>+</sup> *cre* Mutants

Six of the 36 identified single Kan<sup>R</sup> colonies of the third round were chosen for further analysis and referred to as R3M1 to 6 (Round 3 Mutants 1 to 6). Retesting them in the indicator strain revealed that all but one (R3M4) show significant *loxK2* recombination and all are unbiased in their activity on *loxP* (Table 14). Sequencing analysis revealed one amino acid change common to all five mutants having increased *loxK2* activity: a glutamate to glycine exchange at position 262 (E262G) in the J helix of the Cre protein (Figure 13). A second point mutation, a conservative threonine to serine exchange at position 316 (T316S), was identified in three of the mutants with enhanced *loxK2* activity. Two non-conservative mutations (D29A and D189N) were found in two of the five mutants. In addition, ten mutations occurred only once. Therefore, the critical mutation for *loxK2* activity appears to be E262G.

#### Site-Directed Mutagenesis

To address the question of the influence of the different point mutations further, the QuickChange™ Site-Directed Mutagenesis Kit (Stratagene) was applied to generate the following Cre mutants, each confirmed by sequencing: E262G, E262G/D29A, E262G/D189N, and E262G/T316S.

### *In Vivo* Characterization

To elucidate the contribution of specific amino acid changes in conferring altered recombinational specificity to Cre, recombination assays with different *lox* sites were carried out with the following Cre enzymes: the wt enzyme, one of the originally sequenced third round mutants (R3M3), and the generated single and double mutants. All *lox* sites used for the *in vivo* tests were designed to have the same 8 bp spacer region (FAS1) so that recombinational specificity was completely dependent on Cre's recognition of the symmetrical inverted repeats of the *lox* sites. Note that wt Cre-dependent recombination between *loxP* sites bearing the FAS1 spacer does not differ from recombination between original *loxP* sites (Sauer B., *Nucleic Acids Res.*, 24:4608-4613 (1996)).

Figure 14 presents the recombination frequencies of various mutant *lox*<sup>2</sup> substrates and combination of sites (*loxP* with *loxK2* and as control with *loxKI*) from the marker excision assay. Mutant *lox* sites with symmetric nucleotide substitutions at positions 11 and 12 of the *loxP* sequence were tested with the wt enzyme and the five variant Cre mutants, including the multiple mutant R3M3(A). All enzymes showed a maximum in recombination (close to 100%) with thymines at these positions, i.e. with *loxP*<sup>2</sup>. Adenines, i.e. *loxK2*<sup>2</sup>, lowered the recombination frequencies drastically for the wt enzyme, whereas the single and double mutants performed approximately 50% to 70% less effectively. R3M3, however, showed nearly *LoxP*-like activity on the *loxK2*<sup>2</sup> substrate, as seen before with the selection strain (Table 14). Altering the two thymines to guanines resulted in similar recombination frequencies as seen with adenines at these positions. Cytosines, on the other hand, did not result in similar recombination frequencies as seen with thymines, but were surprisingly the least efficiently recognized substitutions of positions 11 and 12. To conclude, it appears that the E262G mutation is necessary and sufficient to significantly increase recombination frequencies on *lox* sequences which are symmetrically altered at positions 11 and 12. Of the additional mutations tested, D29A seems to be slightly beneficial, whereas D189N and T316S appear indifferent or even slightly deleterious for recombination on the variant sites. Thus, additional

5 mutations of R3M3 (Figure 13) must be responsible for its further increased performance on the mutant substrates *in vivo*.

Figure 14B shows the observed recombination frequencies on mixed substrates (e.g. *loxP* with *loxK2*). For both *loxK2* and the control substrate, *loxK1*, recombination with *loxP* by wt enzyme was substantially less than for *loxP*<sup>2</sup> recombination. This recombination frequency was increased dramatically with all of the mutant Cre protein. These results hint that not only a cooperativity in binding of two Cre molecules to one *lox* site exists but moreover also cooperativity between Cre molecules binding to different sites which then are synapsed and recombined. This finding is especially useful for genomic targeting: It suggests that a targeting vector carrying a *loxP* site will be effectively recombined with the endogenous *lox*-like site by the Cre mutants as long as the spacers are compatible.

Noteworthy also is R3M3-Cre's increased ability to recombine *loxK1* by itself, whereas all the mutants, like wt, did not. Again, some additional mutations found in R3M3 (Figure 13) seem to be responsible not only for increased recombination frequencies on *loxK2* but also on *loxK1* compared to Cre bearing the E262G mutation alone or in conjunction with the D29A mutation.

#### 20 *In Vitro* Characterization

*In Vitro* Recombination Assays. For the *in vitro* recombination experiments the same substrates as *in vivo*--that is, pBS632 to pBS641--were used. An *in vitro* recombination experiment using wt, E262G, and R3M3 Cre preps on *loxP*<sup>2</sup>, *loxK2*<sup>2</sup>, and *loxK1*<sup>2</sup> substrate showed that, as seen *in vivo* before, wt Cre is capable to recombine *loxP*<sup>2</sup> substrates only. No recombination products are visible when *loxK2*<sup>2</sup> or *loxK1*<sup>2</sup> substrates were used. By introducing the E262G mutation into Cre, however, recombination of *loxK2*<sup>2</sup> substrates becomes possible at elevated frequencies, and even for *loxK1*<sup>2</sup> substrate recombination products become weakly visible *in vitro*. E262G-Cre activity on the *loxK1* control site *in vitro* but not *in vivo* probably derives from differences in ionic strength and/or enzyme concentration between the assays. Finally, with R3M3 Cre the ability for both *loxK2*<sup>2</sup> and *loxK1*<sup>2</sup> recombination is

5 further increased, as expected from the *in vivo* assays. As mentioned in the *in vivo* results before, guanines at positions 11 and 12 of the *lox* sequence were recognized at similar frequencies as seen with adenines (i.e. *loxK2*), whereas  
10 cytosines were clearly less tolerated by all Cre mutants. Slight differences  
5 between *in vivo* and *in vitro* recombination frequencies are probably due to differences in ionic strength, temperature, DNA condensation, enzyme concentration, etc. In general, the pattern of the *in vitro* recombination  
15 frequencies of the different Cre enzymes on the different *lox* sites mirrors the one seen *in vivo*.

10 *Gelshift Experiments.* Gelshift experiments were applied to address the question of *in vitro* DNA-affinities of the different Cre mutants. As expected  
20 from the previous results, all three Cre enzymes bind with similar efficiency to *loxP*, whereas to *loxK2*, only E262G and R3M3 show binding affinity. Surprisingly, R3M3 binding appears less efficient on the *loxK2* half-site than  
25 binding of E262G, and on the *loxK1* half-site only E262G does bind weakly whereas R3M3 does not.

## DISCUSSION

### *lox sites*

30 *loxP* and *loxK2* differ at several locations as illustrated in Figure 1:  
20 First, the three outermost bp of the inverted repeats are altered (positions 15, 16, 17, 15', 16' and 17'). Second, the entire non-canonical 8 bp spacer is  
35 completely exchanged (positions 4 to 4'), and third, two transversion events (thymine to adenine) are introduced at positions 11 and 12 (and mirrored at 11' and 12') of the *lox* site, mimicking potential recombination targets in  
25 eukaryotes. Only the two mutations at positions 11 and 12 are considered important for inhibiting wt Cre from recognizing the site. Thus, they were the  
40 only alterations in the *loxi* sites used for the *in vivo* and *in vitro* characterization experiments. Other investigations have shown that the two alterations of *loxP* mentioned first are without inhibiting effect on wt Cre (Sauer, B., *Mol. Cell.*  
45 *Biol.*, 7:2087-2096 (1987) and Sauer B., *Nucleic Acids Res.*, 24:4608-4613 (1996)). Noteworthy is however that the altered 8 bp spacer region (FAS1  
30 spacer) does not allow *loxP* - *loxK2* recombination, since the regions where the

5 single-strand cleavages and exchanges take place are not compatible. To allow  
simultaneous monitoring of Cre-mediated recombination both at the wt *loxP*<sup>2</sup>  
and at a mutant *lox*<sup>2</sup> substrate (*loxK2*<sup>2</sup>), incompatible spacer elements were used  
10 to prevent recombination between the two types of *lox* sites by a candidate Cre  
5 mutant with altered specificity. Such recombination might easily compromise  
ready detection of desired Cre specificity mutants. Incompatible spacers  
(original *loxP* and FAS1) formed the basis for the simultaneous selection for  
15 *loxK2* recombination and screen for *loxP* recombination with *E. coli* strain  
BS1494 which led to the disclosed variant Cre recombinases.

10 *LoxKI*, the other *lox* sequence used in this study, bears two critical bp  
exchanges per arm as well, however at different positions: 10 and 14. It was  
20 used as a control *lox* site, addressing the question whether the generated Cre  
mutants with novel specificity for *loxK2* can also tolerate adjacent alterations  
within the *lox* sequence.

#### 15 Cre Mutants after Three Rounds of Directed Evolution

25 Three iterations of the *in vitro* evolution procedure were used to identify  
36 candidates which express Cre mutants that could process *loxK2* (based on the  
selection), as well as *loxP* (based on the simultaneous screen). wt Cre, on the  
30 other hand, only shows marginal activity on *loxK2* when expressed at very high  
20 levels.

#### Characterization

35 Sequence analysis revealed that five mutants with significant *loxK2*  
activity had in common only one point mutation, leading to the amino acid  
change E262G. However, several other point mutations occurred independently  
25 twice or thrice, among which D29A, D189N, and T316S. To investigate the  
influence of the mentioned mutations on the observed phenotype, specific single  
40 and double mutants were generated by site-directed mutagenesis. *In vivo* and *in vitro*  
assays were then carried out with wt-Cre and five different mutants  
(R3M3, E262G, E262G/D29A, E262G/D189N, E262G/T316S) to compare  
45 30 their performance on a variety of different alterations of the *loxK2* site.

The *in vivo* and *in vitro* recombination assays showed a similar pattern  
in recombination frequencies for the different enzymes on the different sites

5 tested. In general, recombination frequencies on mutant substrates were  
distinctively the highest with R3M3. The double mutant E262G/D29A was  
about half as effective as R3M3, whereas the other double mutants and the  
10 single mutant E262G showed slightly further decreased recombination  
5 frequencies on the altered sites. The wt enzyme did not recombine any of the  
mutant substrates presented here, neither *in vivo* nor *in vitro*. With previous  
results showing that single D29A, D189N, and T316S mutants of Cre perform  
15 like the wt enzyme on *loxK2* and *loxK1* *in vivo* and the fact that the E262G  
point-mutation was the only one found independently in all the originally  
10 isolated Cre mutants with *loxK2* specificity, it is clear that E262G is a critical  
mutation that allows Cre to recognize *loxK2*. However, in combination with  
20 E262G, D29A permits still higher recombination frequencies on *lox* sites altered  
at positions 11 and 12. Since R3M3 shows even further increased  
recombination frequencies on the altered sites compared to E262G/D29A, other  
25 of the point mutations identified in this Cre mutant (see Figure 14) must account  
for this increase in activity. Because of its location close to the amino-terminus  
of the J helix, the Q255R mutation of R3M3 could be expected to influence  
DNA contacts. Other mutations may influence protein folding or protein-  
30 protein interactions which could result in a higher flexibility within the Cre-*lox*  
20 interface and thus allowing a better tolerance of alterations of the *lox* sequence.  
This hypothesis is also supported by the observation that R3M3 recognizes the  
*loxK1* site at frequencies similar to E262G recognizing *loxK2*. The double and  
35 single Cre mutants, on the other hand, did not show activity on *loxK1*. In  
addition, the gel-shift experiments showed that R3M3-Cre's binding affinity for  
25 *loxK2* and *loxK1* half-sites is less than E262G-Cre's and the three double  
40 mutant's. Taking these results together, other mutations of R3M3 must further  
influence Cre-*lox* interactions to allow enhanced recombination on *loxK2*. On  
the one hand, this results in less efficient binding to *lox* half-sites, on the other  
45 hand, when complete *lox* sites are available, the cooperativity phenomenon  
30 between Cre enzymes binding to the same and different *lox* sites may  
compensate for this loss in binding activity. Then the postulated increased  
flexibility between DNA and protein seems to become advantageous for

5 recognizing and recombining altered *lox* sites, as seen with the *in vivo* and *in vitro* recombination assays.

Alternatively, some mutations, also silent ones, could influence protein  
expression, leading to a faster accumulation of enzymes and consequently to  
10 5 higher recombination frequency. An *E. coli* codon usage table suggested,  
however, that none of the identified mutations should improve Cre expression  
in *E. coli* remarkably.

#### 15 Modeling

With the published crystal structure of four Cre molecules bound to two  
10 synapsed *loxA* sites after the first single strand cleavage (Guo et al., *Nature*,  
389:40-46 (1997)) the identified point mutations were analyzed for being  
20 involved in DNA and/or protein interactions. All of them, including the E262  
position in the J-helix of Cre were found to be not involved in either interactions  
in this state of the reaction. This observation indicates that the mutations which  
25 were found to account for the described novel substrate recognition *in vivo* and  
*in vitro* lead to this new phenotype in a less direct and obvious manner. As  
mentioned earlier, they may influence protein folding, resulting in a higher  
flexibility within the Cre-*lox* interface. This hypothesis is especially well  
30 supported with the described differences between binding to *lox* halvesites and  
20 recombination. Alternatively, they may still be involved in protein-protein  
(D29A) or protein-DNA (E262G) contacts before or after the formation of the  
clamp-like strand-exchange state.

#### 35 *In Vivo* and *In Vitro* Recombination

Some additional variations of the *loxK2* site were tested also. The list in  
25 Table 17 shows all the sequences of *lox* sites which were tested and assigns  
them a name, as well as their plasmid (pBS) number. In Figure 15 the obtained  
40 *in vivo* recombination frequencies on all the variants of *loxK2* are indicated.  
The additional results indicate that alterations at position 12 of the *lox* halvesite  
are of more importance for Cre-based recombination than ones at position 11.

45 30 The *in vitro* recombination frequencies of all six Cre enzymes tested on  
the *lox* sites listed above are given in table 15. The frequencies were calculated  
after quantitation of the brightness of fluorescence of the Ethidium-Bromide-

5 stained DNA fragments on agarose-gels. Differences in temperature, ionic  
strength, medium composition, and enzyme concentration probably account for  
the observed differences between *in vivo* and *in vitro* recombination results.  
10 Most strikingly, *loxP* is no longer recognized with the highest frequencies *in*  
5 *vitro*. However, when the ionic strength in the *in vitro* assay was increased  
results began to resemble the ones seen *in vivo*. Thus, efficiency of  
recombination with variant *lox* sites by each of the Cre mutant and the wt  
15 enzyme can be further controlled *in vitro* by adjusting ionic strength and other  
*in vitro* conditions.

10 Qualitatively, however, the *in vivo* and *in vitro* recombination  
frequencies mirror each other. These novel Cre mutants thus possess a  
20 specificity for substrates (*loxK2* and its derivatives) which are not recognized by  
the wt enzyme.

#### Gelshift Assays

15 DNA binding (gelshift) experiments were also done with the generated  
Cre double-mutants (E262G with D29A, D189N, or T316S) to analyzing  
binding of the variant Cre recombinases to various recombination sites. In table  
16 the observed mean percentages of binding to *loxP*, *loxK2*, and *loxK1*  
30 halfsites with the five different Cre mutants and the wt enzyme are given. As  
shown, all mutants - in contrast to the wt enzyme - do bind to *loxK2* with  
20 similar frequencies, except R3M3 which shows surprisingly low retardation. As  
discussed above, this phenomenon may be explained with an increased  
35 tolerance, i.e., flexibility, of R3M3 for altered *lox* sites. On halfsites which  
precludes the cooperativity between Cre molecules in binding R3M3 cannot  
25 bind as tight as wt or the single and double mutants. With the *loxK1* halfsite  
this 'binding versus recombination' difference is even more strikingly.  
40 Whereas wt and the R3M3 mutant cannot bind to the halfsite, all the other  
mutants can. Yet, recombination of *loxK1* substrates was seen with R3M3 Cre,  
45 only. These results show clearly that simple DNA affinity does not correlate in  
30 a one-to-one fashion with recombination. Thus, inappropriate DNA binding by  
recombinases likely can lead to a block in recombination.

**Conclusions** The *in vivo* data show that:

1. The E262G mutation confers a generally elevated level of recombination at a number of variant *lox* sites, those having any of a large number of alterations at positions 11 and 12 (and mirroring 11' and 12' alterations).

2. The D189N mutation in conjunction with the E262G mutation appears to fine tune the broadened specificity of the E262G mutation by reducing recombination at the *loxK2* variants 'GG', 'TC' and 'CC' without decreasing recombination at *loxK2* and the 'GT' variant. This mutation is thus useful to limit the broadened specificity of E262G.

3. The T316S mutation when in conjunction with E262G provides a slight boost in recombination frequency with the *loxK2* variants 'GT' and 'CC', and has no deleterious effect on recombination at the other variant *loxK2* sites.

4. The D29A mutation together with E262G boosts recombination at *loxK2* and the variants 'CC' and 'GG'. D29A does not reduce recombination at the other variant *lox* sites or at *loxP*.

5. Some of the additional mutations of R3M3 must account for the further increased recombination frequency on any of the tested mutant *lox* sites, including *loxK1* but do not compromise *loxP* recognition.

The disclosed variant recombinases have a number of useful features and applications. By recognizing an altered, user-defined target site, they were designed to allow both genetic targeting events in prokaryotes and eukaryotes like wt Cre but on different sites and *in vitro* recombination strategies. With a wider variety of possible target sequences being now accessible, multiple and defined genomic alterations will now become feasible. This opens more possibilities in designing genomic manipulations in all DNA-based organisms by site-specific recombination.

With the various genome projects under way today, there will be more and more applications for site-specific recombination to study the impact of genes or genetic control elements by genomic engineering. In addition, genome manipulations are also more frequently used to express recombinant proteins within the organism of choice.

**Table 13.** Presented are the results of the first four rounds of the described mutagenesis procedure for *loxK2* specificity. Column one indicates the round of the mutagenesis procedure (round 0 indicates that no mutagenesis event has taken place yet, wt enzyme, only) and column two the allotted time for *cre* expression in induction medium. Column three presents the observed frequencies of Kan<sup>R</sup>, indicating *loxK2* recombination by a mutant candidate. Of note, all Kan<sup>R</sup> colonies found during this experiment were white, indicating effective *loxP* recombination. In column four the actual number of Kan<sup>R</sup> colonies found in each round of 1000 to 10<sup>4</sup> plated is given, and the last column presents the frequency of white colonies on non-selection plates which is decreasing with every round.

**Table 14.** Presented are the frequencies of Kan<sup>R</sup> of six chosen candidate mutants from round three of the mutagenesis procedure (R3M1 to 6) when retesting them in the indicator strain. Also, they were tested for their performance on *loxP* sites in identical fashion. The obtained Kan<sup>R</sup> frequencies indicate the percentages of recombination within the allotted induction time of 2.5 hours.

Table 2: List of Strains Used for Selection and Screen

BS #	Description	Substrates	Selection and Screen
1493	BS583 [pBS583]	<i>loxK1<sup>2</sup>/loxP<sup>2</sup></i>	<i>neo</i> selection and <i>lacZ</i> screen, respectively
1494	BS583 [pBS584]	<i>loxK2<sup>2</sup>/loxP<sup>2</sup></i>	<i>neo</i> selection and <i>lacZ</i> screen, respectively
1541	BS583 [pBS613]	<i>loxP<sup>2</sup>/loxP<sup>2</sup></i>	<i>neo</i> selection and <i>lacZ</i> screen, respectively
1523	NS2300 [pBS601]	<i>loxK1<sup>2</sup>/loxP<sup>2</sup></i>	<i>abfA.st</i> screen and <i>neo</i> selection, respectively
1524	NS2300 [pBS602]	<i>loxK2<sup>2</sup>/loxP<sup>2</sup></i>	<i>abfA.st</i> screen and <i>neo</i> selection, respectively

List of the *E. coli* strains used for the described selection and screening procedures for the desired Cre mutants. BS1493 and BS1494 served as selection strains for K1<sup>+</sup> or K2<sup>+</sup> Cre mutants, respectively. BS1541 was used as a control strain to determine *loxP* activity. BS1523 and BS1524 were used to select for P<sup>+</sup> mutants. Construction and application of the different selection procedures are explained in detail in the text.

Table 3: Sequencing Primers for *cre*

Primer BSB#	Sequence 5' - 3'	Position (5', 3') in <i>cre</i>
454	TTT GGG CTA GCG AAT TCG AG	-55, -36
455	TTT GGG CCA GCT AAA CAT GC	273, 292
456	CGG TGG GAG AAT GTT AAT CC	567, 586
457	GGA CAC AGT GCC CGT GTC	862, 879
458	TCT GCG TTC TGA TTT AAT CTG	1117, 1097
459	CCA GGC CAG GTA TCT CTG	858, 841
460	GTA CGT GAG ATA TCT TTA ACC C	563, 542
461	TTG CTG GAT AGT TTT TAC TGC C	270, 249

Presented are the eight sequencing primers used for sequencing of the entire *cre* gene in both directions and their positions within the *cre* coding sequence. Primers BSB 454 to BSB457 allow sequencing of the coding strand, whereas the four remaining primers have been designed for the non-coding strand. In order to achieve a complete sequence information in both directions, about 350 bases have to be read out of each sequencing reaction. The positions given for each primer refer to its 5' and 3' end, with position 1 referring to the adenine of the ATG start codon of *cre*.

Table 4: Primers Used in the QuickChange™ Site-Directed Mutagenesis Kit (Stratagene)

Primer BSB#	Sequence 5' - 3'	Position (5', 3') in <i>cre</i>
465	GCTATCAACTCGGCCCTGGGAGGGATTTTGAAGCAACTCATCG	765, 809
466	GAGTTGCTTCAAAAATCCCTCCAGGGCGCGAGTTGATAGCTGGC	805, 761
467	GCTATCAACTCGGCCCTGGCAGGGATTTTGAAGCAACTCATCG	765, 809
468	GAGTTGCTTCAAAAATCCCTGCCAGGGCGCGAGTTGATAGCTGGC	805, 761
469	GCTATCAACTCGGCCNNNNNNNNATTTTGAAGCAACTCATCG	765, 809
470	GAGTTGCTTCAAAAATNNNNNNNNNGGCGCGAGTTGATAGCTGGC	805, 761

Illustrated are the mutant oligonucleotides used in the QuickChange™ Site-Directed Mutagenesis Kit (Stratagene) to introduce single point mutations (mismatches with the wt sequence are highlighted by bold letters). With BSB465/466 a E262G mutation, and with BSB467/468 a E262A mutation is introduced in Cre. The last set of primers (BSB469/470), as indicated by the bold N, represent an equimolar mixture of all possible bases at the assigned positions. This mixture results in all possible aa combinations at positions 261 to 263 of Cre. As mentioned before, the indicated positions refer to 5' and 3' end of the oligonucleotides in the *cre* coding sequence.

Table 5: Comparison of Mutagenic and Non-Mutagenic PCR on wt *cre*

Induction of <i>cre</i> Expression	Non-Mutagenic PCR		Mutagenic PCR	
	Blue	White	Blue	White
SOC, 2.5h	> 99.6%	< 0.4%	> 99.5%	< 0.5%
SOB+ara, 2.5h	19.7%	81.3%	60.8%	39.2%

The experiments presented here served two purposes: (i) to compare the influence of glucose (SOC medium) and arabinose on *cre* expression with pBAD33, (ii) to compare the frequencies of deleterious point mutations occurring during non-mutagenic and error-prone PCR. The first column indicates the medium used for repression (SOC), or induction of *cre* expression (SOB+ara). The second column compares the observed frequencies of blue and white colonies after a non-mutagenic PCR, the third column after a mutagenic PCR. Blue colonies indicate, that the *loxP*<sup>2</sup> flanked *lacZ* gene of BS583 is not excised by Cre, white ones its Cre-mediated excision. Note that the number of blue colonies resulting from empty pBAD33 vector (approximately 1% in all experiments) was subtracted before calculating the presented values. This blue background was determined with a control ligation lacking *cre* insertion. With glucose in the medium, only blue colonies are found, indicating that *cre* is not expressed. With arabinose induction, however, the frequencies of disabled Cre mutants (blue) can be determined, increasing three times under mutagenic conditions.

Table 6: The First Four Rounds on *loxK2*<sup>2</sup>

Round	Status	Induction	% White (Non-Selection)	% Kan <sup>R</sup> (White)	Nb. Kan <sup>R</sup> (White)	Nb. Pooled for the Next Round
1	ox	4 h	83.9%	0.01%	1	none
	mx	4 h	42.4%	0.16%	6	6
2	mxox	2.5 h	88.3%	0.02%	1	none
		4 h	82.5%	0.56%	47	47
	mxmx	2.5 h	13%	< 0.01%	0	none
		4 h	24%	0.3%	5	none
3	mxoxox	2.5 h	80.7%	0.2%	36	30
		4 h	85%	25%	10 <sup>3</sup>	none
	mxoxmx	2.5 h	25.7%	0.02%	1	none
		4 h	21.8%	5%	250	none
4	mxoxoxox	2.5 h	75%	4.6%	102	none
	mxoxoxmx	2.5 h	< 7.7%	0.3%	3	none

Presented are the results of the first four rounds of the described mutagenesis procedure for *loxK2* specificity. The following symbols for the status of the *cre* pools have been used: "o" indicates a non-mutagenic, "m" a mutagenic PCR, and "x" stands for the shuffling step by DNase I digest and reassembly. Column one indicates the round of the mutagenesis procedure, column two its status as mentioned above, and column three the time permitted for *cre* expression in induction medium. In column four, the calculated percentages of white colonies on non-selection plates is indicated, decreasing with every round, especially when the pools were subjected to mutagenic PCR's twice. Column five presents the observed frequencies of Kan<sup>R</sup>, indicating *loxK2* recombination by a mutant candidate. Of note, that all Kan<sup>R</sup> colonies found during this experiment were white, indicating effective *loxP* recombination. The two last columns give the actual numbers of Kan<sup>R</sup> colonies found in each round (1/5 dilution) and mark the ones used as starting material for the next round.

Table 7: Functional Evaluation of Six Selected mxoxox K2\* Mutants and wt Cre

Cre Candidate	Induction Time	% Kan <sup>R</sup> with		
		BS1494 ( <i>loxK2</i> <sup>2</sup> )	BS1541 ( <i>loxP</i> <sup>2</sup> )	BS1493 ( <i>loxK1</i> <sup>2</sup> )
wt	2.5 h	0.002%	93.9%	0.001%
	4 h	n.d.	n.d.	0.002%
mxoxox 1	2.5 h	66.4%	98.2%	0.004%
	4 h	n.d.	n.d.	0.005%
mxoxox 2	2.5 h	53.6%	98%	0.007%
	4 h	n.d.	n.d.	0.005%
mxoxox 3	2.5 h	66.5%	80.5%	0.007%
	4 h	n.d.	n.d.	0.01%
mxoxox 4	2.5 h	0.05%	80.3%	0.008%
	4 h	n.d.	n.d.	0.006%
mxoxox 5	2.5 h	15.2%	81.2%	0.006%
	4 h	n.d.	n.d.	0.009%
mxoxox 6	2.5 h	3.1%	83.7%	0.003%
	4 h	n.d.	n.d.	0.01%

The results of functional testing of the six chosen K2\*/P\* Cre candidates compared to wt Cre are listed. All enzymes were checked on *loxK2*, *loxP* and *loxK1* recombination in the selection strains BS1494, BS1493 and the control strain BS1541. In the first column the Cre candidate is indicated, in the second the allowed time for cre expression. The three following columns show the calculated percentages of Kan<sup>R</sup> when subjecting the different Cre mutants and wt Cre to the three selection strains. The frequency of the observed Kan<sup>R</sup> phenotype is considered as an indicator for the frequency of *lox* recombination: On *loxK2*, all mutants, except mxoxox 4, show remarkably increased Kan<sup>R</sup> frequencies compared to wt Cre. On *loxP*, all (including mxoxox 4), show very similar results as wt Cre, indicating, that *loxP* recombination is at best slightly affected. On *loxK1* neither mutant nor wt Cre show more than background activity, even after four hours of cre expression. The obtained mutants are therefore characterized by a broader substrate recognition compared to the wt enzyme.

Table 8: Sequence Analysis of the Six mxoxox K2\*/P\* Mutants

Cre Candidate	Codon Position	Change in <i>cre</i> wt → mut		aa Change	Nb. of Isolates	Position in Cre
mxoxox 1	45	CCG	CCT	(P15P)	2	(N-terminal)
	86	GAC	GCC	D29A	2	A
	565	GAC	AAC	D189N	2	loop 1 - 2
	642	AGC	AGT	(S214S)	1	(N-term. of I)
	679	GTC	ATC	V227I	1	I
	785	GAA	GGA	E262G*	5	J
mxoxox 2	45	CCG	CCT	(P15P)	2	(N-terminal)
	413	GAA	GGA	E138G	1	N-term. of F
	760	AGC	GGC	S254G	1	loop 5 - J
	785	GAA	GGA	E262G*	5	J
	861	TCT	TCC	(S287S*)	2	N-term. of K
	946	ACC	TCC	T316S	3	loop L - M
mxoxox 3	46	GTC	ATC	V16I	1	N-terminal
	565	GAC	AAC	D189N	2	loop 1 - 2
	592	GGC	AGC	G198S	1	C-term. of 2
	668	CQA	CAA	R223Q	1	I
	764	CAG	CGG	Q255R	1	loop 5 - J
	785	GAA	GGA	E262G*	5	J
	861	TCT	TCC	(S287S*)	2	N-term. of K
	920	CCG	CTG	P307L	1	L
mxoxox 4	230	TAT	TGT	Y77C	1	C
	412	GAA	AAA	E138K	1	N-term. of F
	851	CTG	CAG	L284Q	2	loop J - K
mxoxox 5	86	GAC	GCC	D29A	2	A
	302	CGG	CAG	R101Q*	1	D
	659	CTG	CAG	L220Q	1	I
	785	GAA	GGA	E262G*	5	J
	946	ACC	TCC	T316S	3	loop L - M
mxoxox 6	785	GAA	GGA	E262G*	5	J
	851	CTG	CAG	L284Q	2	loop J - K
	946	ACC	TCC	T316S	3	loop L - M

The point mutations identified by sequence analysis in the six selected K2\*/P\* mutants are listed. Column one identifies the mutants, column two the observed point mutations and their position in the *cre* coding sequence. Column three indicates the resulting aa changes and their positions in the Cre enzyme (silent mutations in parenthesis). DNA contacting residues, according to the crystal structure (Guo, *et al.*, 1997), are marked by an asterisk. Column four indicates how often the different point mutations were independently found in the pool of the six mutants, and column five shows where the aa changes are located in the secondary structure of the protein. Of the altogether 31 point mutations listed, five are silent. Only one aa change, E262G, is common to all mutants with remarkably increased *loxK2* activity (mxoxox 1, 2, 3, 5, and 6), suggesting that this mutation is the essential one for the observed phenotype. Nine point mutations occurred twice or three times in the pool, whereas eleven were found in one mutant only. For further information, see text and appendices A, B, and C.

Table 9: Functional Evaluation of Three Defined Cre Mutants for Position 262 and wt Cre

Cre Mutant	Induction Time	% Kan <sup>R</sup> with		
		BS1494 ( <i>loxK2</i> <sup>2</sup> )	BS1541 ( <i>loxP</i> <sup>2</sup> )	BS1493 ( <i>loxK1</i> <sup>2</sup> )
wt	2.5 h	0.002%	93.9%	0.001%
E262G	2.5 h	2.2%	84.2%	0.004%
E262A	2.5 h	0.4%	82.3%	0.006%
E262W	2.5 h	2.9%	81.7%	0.003%

The results of functional testing of three defined Cre mutants for position 262 compared to wt Cre are listed. All enzymes were checked on *loxK2*, *loxP* and *loxK1* recombination in the selection strains BS1494, BS1493 and the control strain BS1541. In the first column the Cre candidate is indicated, in the second the allowed time for cre expression. The three following columns show the calculated percentages of Kan<sup>R</sup> when subjecting the different Cre mutants and wt Cre to the three selection strains. The frequency of the observed Kan<sup>R</sup> phenotype is considered as an indicator for the frequency of *lox*<sup>2</sup> recombination: On *loxK2*, E262G and E262W show approximately an 10<sup>3</sup> fold increased Kan<sup>R</sup> frequency compared to wt Cre, whereas E262A results in a 200 fold increase. On *loxP*, all show very similar results as wt Cre, indicating, that *loxP* recombination is at best slightly affected. On *loxK1* neither mutant nor wt Cre show more than background activity. This indicates, that E262G and E262W alone are sufficient to remarkably increase Cre's activity on *loxK2*, that these mutations do not increase *loxK1* recognition, and that *loxP* activity may be slightly affected.

Table 10

*In Vitro Evolution of the Cre recombinase*

wtCre	TTTGGGCTAG	CGAATTCGAG	CTCGGTACCC	GGGGATCCTC	TAGACTGAGT
mx000x1	TTTGGGCTAG	CGAATTCGAG	CTCGGTACCC	GGGGATCCTC	TAGACTGAGT
mx000x2	TTTGGGCTAG	CGAATTCGAG	CTCGGTACCC	GGGGATCCTC	TAGACTGAGT
mx000x3	TTTGGGCTAG	CGAATTCGAG	CTCGGTACCC	GGGGATCCTC	TAGACTGAGT
mx000x4	TTTGGGCTAG	CGAATTCGAG	CTCGGTACCC	GGGGATCCTC	TAGACTGAGT
mx000x5	TTTGGGCTAG	CGAATTCGAG	CTCGGTACCC	GGGGATCCTC	TAGACTGAGT
mx000x6	TTTGGGCTAG	CGAATTCGAG	CTCGGTACCC	GGGGATCCTC	TAGACTGAGT
-55					
wtCre	GTGAAATCTC	CAATTTACTG	ACCGTACACC	AAAAATTGCC	TGCATTACCG
mx000x1	GTGAAATCTC	CAATTTACTG	ACCGTACACC	AAAAATTGCC	TGCATTACCT
mx000x2	GTGAAATCTC	CAATTTACTG	ACCGTACACC	AAAAATTGCC	TGCATTACCT
mx000x3	GTGAAATCTC	CAATTTACTG	ACCGTACACC	AAAAATTGCC	TGCATTACCG
mx000x4	GTGAAATCTC	CAATTTACTG	ACCGTACACC	AAAAATTGCC	TGCATTACCG
mx000x5	GTGAAATCTC	CAATTTACTG	ACCGTACACC	AAAAATTGCC	TGCATTACCG
mx000x6	GTGAAATCTC	CAATTTACTG	ACCGTACACC	AAAAATTGCC	TGCATTACCG
-5					
wtCre	GTCGATGCAA	CGAGTGATGA	GGTTGGCAAG	AACCTGATGG	ACATGTTICAG
mx000x1	GTCGATGCAA	CGAGTGATGA	GGTTGGCAAG	AACCTGATGG	CCATGTTICAG
mx000x2	GTCGATGCAA	CGAGTGATGA	GGTTGGCAAG	AACCTGATGG	ACATGTTICAG
mx000x3	ATCGATGCAA	CGAGTGATGA	GGTTGGCAAG	AACCTGATGG	ACATGTTICAG
mx000x4	GTCGATGCAA	CGAGTGATGA	GGTTGGCAAG	AACCTGATGG	ACATGTTICAG
mx000x5	GTCGATGCAA	CGAGTGATGA	GGTTGGCAAG	AACCTGATGG	CCATGTTICAG
mx000x6	GTCGATGCAA	CGAGTGATGA	GGTTGGCAAG	AACCTGATGG	ACATGTTICAG
46					
wtCre	GGATCGCCAG	GCGTTTTCIG	AGCATAOCTG	GAAAATGCCT	CTGTCCGTTT
mx000x1	GGATCGCCAG	GCGTTTTCIG	AGCATAOCTG	GAAAATGCCT	CTGTCCGTTT
mx000x2	GGATCGCCAG	GCGTTTTCIG	AGCATAOCTG	GAAAATGCCT	CTGTCCGTTT
mx000x3	GGATCGCCAG	GCGTTTTCIG	AGCATAOCTG	GAAAATGCCT	CTGTCCGTTT
mx000x4	GGATCGCCAG	GCGTTTTCIG	AGCATAOCTG	GAAAATGCCT	CTGTCCGTTT
mx000x5	GGATCGCCAG	GCGTTTTCIG	AGCATAOCTG	GAAAATGCCT	CTGTCCGTTT
mx000x6	GGATCGCCAG	GCGTTTTCIG	AGCATAOCTG	GAAAATGCCT	CTGTCCGTTT
96					
wtCre	GCGGCTGGTG	GCGGCGCATGG	TGCAAGTTGA	ATAACCGGAA	ATGGTTTCCC
mx000x1	GCGGCTGGTG	GCGGCGCATGG	TGCAAGTTGA	ATAACCGGAA	ATGGTTTCCC
mx000x2	GCGGCTGGTG	GCGGCGCATGG	TGCAAGTTGA	ATAACCGGAA	ATGGTTTCCC
mx000x3	GCGGCTGGTG	GCGGCGCATGG	TGCAAGTTGA	ATAACCGGAA	ATGGTTTCCC
mx000x4	GCGGCTGGTG	GCGGCGCATGG	TGCAAGTTGA	ATAACCGGAA	ATGGTTTCCC
mx000x5	GCGGCTGGTG	GCGGCGCATGG	TGCAAGTTGA	ATAACCGGAA	ATGGTTTCCC
mx000x6	GCGGCTGGTG	GCGGCGCATGG	TGCAAGTTGA	ATAACCGGAA	ATGGTTTCCC
146					
wtCre	GCAGAACCTG	AAGATGTTCC	CGATTATCTT	CTATATCTTC	AGGCGCGCGG
mx000x1	GCAGAACCTG	AAGATGTTCC	CGATTATCTT	CTATATCTTC	AGGCGCGCGG
mx000x2	GCAGAACCTG	AAGATGTTCC	CGATTATCTT	CTATATCTTC	AGGCGCGCGG
mx000x3	GCAGAACCTG	AAGATGTTCC	CGATTATCTT	CTATATCTTC	AGGCGCGCGG
mx000x4	GCAGAACCTG	AAGATGTTCC	CGATTATCTT	CTATATCTTC	AGGCGCGCGG
mx000x5	GCAGAACCTG	AAGATGTTCC	CGATTATCTT	CTATATCTTC	AGGCGCGCGG
mx000x6	GCAGAACCTG	AAGATGTTCC	CGATTATCTT	CTATATCTTC	AGGCGCGCGG
196					

Table 10 (cont.)

*In Vitro Evolution of the Cre recombinase*

wtCre	TCTGGCAGTA	AAAACATGCC	AGCAACATTT	GGGCCAGCTA	AACATGCTTC
mxoxox1	TCTGGCAGTA	AAAACATGCC	AGCAACATTT	GGGCCAGCTA	AACATGCTTC
mxoxox2	TCTGGCAGTA	AAAACATGCC	AGCAACATTT	GGGCCAGCTA	AACATGCTTC
mxoxox3	TCTGGCAGTA	AAAACATGCC	AGCAACATTT	GGGCCAGCTA	AACATGCTTC
mxoxox4	TCTGGCAGTA	AAAACATGCC	AGCAACATTT	GGGCCAGCTA	AACATGCTTC
mxoxox5	TCTGGCAGTA	AAAACATGCC	AGCAACATTT	GGGCCAGCTA	AACATGCTTC
mxoxox6	TCTGGCAGTA	AAAACATGCC	AGCAACATTT	GGGCCAGCTA	AACATGCTTC
	246				
wtCre	ATCGTCGGTC	CGGGCTGCCA	CGACCAAGTG	ACAGCAATGC	TGTTTCACTG
mxoxox1	ATCGTCGGTC	CGGGCTGCCA	CGACCAAGTG	ACAGCAATGC	TGTTTCACTG
mxoxox2	ATCGTCGGTC	CGGGCTGCCA	CGACCAAGTG	ACAGCAATGC	TGTTTCACTG
mxoxox3	ATCGTCGGTC	CGGGCTGCCA	CGACCAAGTG	ACAGCAATGC	TGTTTCACTG
mxoxox4	ATCGTCGGTC	CGGGCTGCCA	CGACCAAGTG	ACAGCAATGC	TGTTTCACTG
mxoxox5	ATCGTCGGTC	CGGGCTGCCA	CGACCAAGTG	ACAGCAATGC	TGTTTCACTG
mxoxox6	ATCGTCGGTC	CGGGCTGCCA	CGACCAAGTG	ACAGCAATGC	TGTTTCACTG
	296				
wtCre	GTTATCGGGC	GGATCGGAAA	AGAAAAAGTT	GATGCCGGTG	AACGTGCAAA
mxoxox1	GTTATCGGGC	GGATCGGAAA	AGAAAAAGTT	GATGCCGGTG	AACGTGCAAA
mxoxox2	GTTATCGGGC	GGATCGGAAA	AGAAAAAGTT	GATGCCGGTG	AACGTGCAAA
mxoxox3	GTTATCGGGC	GGATCGGAAA	AGAAAAAGTT	GATGCCGGTG	AACGTGCAAA
mxoxox4	GTTATCGGGC	GGATCGGAAA	AGAAAAAGTT	GATGCCGGTG	AACGTGCAAA
mxoxox5	GTTATCGGGC	GGATCGGAAA	AGAAAAAGTT	GATGCCGGTG	AACGTGCAAA
mxoxox6	GTTATCGGGC	GGATCGGAAA	AGAAAAAGTT	GATGCCGGTG	AACGTGCAAA
	346				
wtCre	ACAGGCTCTA	CGGTTGGAAC	GCACGTGATT	CGACCAAGTT	CGTTCACTCA
mxoxox1	ACAGGCTCTA	CGGTTGGAAC	GCACGTGATT	CGACCAAGTT	CGTTCACTCA
mxoxox2	ACAGGCTCTA	CGGTTGGAAC	GCACGTGATT	CGACCAAGTT	CGTTCACTCA
mxoxox3	ACAGGCTCTA	CGGTTGGAAC	GCACGTGATT	CGACCAAGTT	CGTTCACTCA
mxoxox4	ACAGGCTCTA	CGGTTGGAAC	GCACGTGATT	CGACCAAGTT	CGTTCACTCA
mxoxox5	ACAGGCTCTA	CGGTTGGAAC	GCACGTGATT	CGACCAAGTT	CGTTCACTCA
mxoxox6	ACAGGCTCTA	CGGTTGGAAC	GCACGTGATT	CGACCAAGTT	CGTTCACTCA
	396				
wtCre	TGGAAAATAG	CGATCGCTGC	CAGGATATAC	GTAATCTGGC	ATTTCGCGG
mxoxox1	TGGAAAATAG	CGATCGCTGC	CAGGATATAC	GTAATCTGGC	ATTTCGCGG
mxoxox2	TGGAAAATAG	CGATCGCTGC	CAGGATATAC	GTAATCTGGC	ATTTCGCGG
mxoxox3	TGGAAAATAG	CGATCGCTGC	CAGGATATAC	GTAATCTGGC	ATTTCGCGG
mxoxox4	TGGAAAATAG	CGATCGCTGC	CAGGATATAC	GTAATCTGGC	ATTTCGCGG
mxoxox5	TGGAAAATAG	CGATCGCTGC	CAGGATATAC	GTAATCTGGC	ATTTCGCGG
mxoxox6	TGGAAAATAG	CGATCGCTGC	CAGGATATAC	GTAATCTGGC	ATTTCGCGG
	446				
wtCre	ATTGCTTATA	ACACCCGTGT	ACGTATAGCC	GAAATTGCCA	GGATCAGGGT
mxoxox1	ATTGCTTATA	ACACCCGTGT	ACGTATAGCC	GAAATTGCCA	GGATCAGGGT
mxoxox2	ATTGCTTATA	ACACCCGTGT	ACGTATAGCC	GAAATTGCCA	GGATCAGGGT
mxoxox3	ATTGCTTATA	ACACCCGTGT	ACGTATAGCC	GAAATTGCCA	GGATCAGGGT
mxoxox4	ATTGCTTATA	ACACCCGTGT	ACGTATAGCC	GAAATTGCCA	GGATCAGGGT
mxoxox5	ATTGCTTATA	ACACCCGTGT	ACGTATAGCC	GAAATTGCCA	GGATCAGGGT
mxoxox6	ATTGCTTATA	ACACCCGTGT	ACGTATAGCC	GAAATTGCCA	GGATCAGGGT
	496				

Table 10 (con<sup>1</sup>)*In Vitro* Evolution of the Cre recombinase

wtCre	TAAAGATATC	TCACGTACTG	ACGGTGGGAG	AATGTTAATC	CATATTGGCA
mxoxox1	TAAAGATATC	TCACGTACTA	ACGGTGGGAG	AATGTTAATC	CATATTGGCA
mxoxox2	TAAAGATATC	TCACGTACTG	ACGGTGGGAG	AATGTTAATC	CATATTGGCA
mxoxox3	TAAAGATATC	TCACGTACTA	ACGGTGGGAG	AATGTTAATC	CATATTGGCA
mxoxox4	TAAAGATATC	TCACGTACTG	ACGGTGGGAG	AATGTTAATC	CATATTGGCA
mxoxox5	TAAAGATATC	TCACGTACTG	ACGGTGGGAG	AATGTTAATC	CATATTGGCA
mxoxox6	TAAAGATATC	TCACGTACTG	ACGGTGGGAG	AATGTTAATC	CATATTGGCA
546					
wtCre	GAACGAAAAC	GCTGGTTAGC	ACCGCAGGTG	TAGAGAAGGC	ACTTAGCCTG
mxoxox1	GAACGAAAAC	GCTGGTTAGC	ACCGCAGGTG	TAGAGAAGGC	ACTTAGCCTG
mxoxox2	GAACGAAAAC	GCTGGTTAGC	ACCGCAGGTG	TAGAGAAGGC	ACTTAGCCTG
mxoxox3	GAACGAAAAC	GCTGGTTAGC	ACCGCAGGTG	TAGAGAAGGC	ACTTAGCCTG
mxoxox4	GAACGAAAAC	GCTGGTTAGC	ACCGCAGGTG	TAGAGAAGGC	ACTTAGCCTG
mxoxox5	GAACGAAAAC	GCTGGTTAGC	ACCGCAGGTG	TAGAGAAGGC	ACTTAGCCTG
mxoxox6	GAACGAAAAC	GCTGGTTAGC	ACCGCAGGTG	TAGAGAAGGC	ACTTAGCCTG
596					
wtCre	GCGGTAACCTA	AACCTGGTGA	GCGATGGATT	TCCGTCTCTG	GTGTAGCTGA
mxoxox1	GCGGTAACCTA	AACCTGGTGA	GCGATGGATT	TCCGTCTCTG	GTGTAGCTGA
mxoxox2	GCGGTAACCTA	AACCTGGTGA	GCGATGGATT	TCCGTCTCTG	GTGTAGCTGA
mxoxox3	GCGGTAACCTA	AACCTGGTGA	GCGATGGATT	TCCGTCTCTG	GTGTAGCTGA
mxoxox4	GCGGTAACCTA	AACCTGGTGA	GCGATGGATT	TCCGTCTCTG	GTGTAGCTGA
mxoxox5	GCGGTAACCTA	AACCTGGTGA	GCGATGGATT	TCCGTCTCTG	GTGTAGCTGA
mxoxox6	GCGGTAACCTA	AACCTGGTGA	GCGATGGATT	TCCGTCTCTG	GTGTAGCTGA
646					
wtCre	TGATCCGAAT	AACCTACCTGT	TTTGCCGGGT	CAGAAAAAAT	GGTGTGCCG
mxoxox1	TGATCCGAAT	AACCTACCTGT	TTTGCCGGGT	CAGAAAAAAT	GGTGTGCCG
mxoxox2	TGATCCGAAT	AACCTACCTGT	TTTGCCGGGT	CAGAAAAAAT	GGTGTGCCG
mxoxox3	TGATCCGAAT	AACCTACCTGT	TTTGCCGGGT	CAGAAAAAAT	GGTGTGCCG
mxoxox4	TGATCCGAAT	AACCTACCTGT	TTTGCCGGGT	CAGAAAAAAT	GGTGTGCCG
mxoxox5	TGATCCGAAT	AACCTACCTGT	TTTGCCGGGT	CAGAAAAAAT	GGTGTGCCG
mxoxox6	TGATCCGAAT	AACCTACCTGT	TTTGCCGGGT	CAGAAAAAAT	GGTGTGCCG
696					
wtCre	CGCCATCTGC	CACCCAGCCAG	CTATCAACTC	CGCCCTGGG	AGGGATTTTT
mxoxox1	CGCCATCTGC	CACCCAGCCAG	CTATCAACTC	CGCCCTGGG	AGGGATTTTT
mxoxox2	CGCCATCTGC	CACCCAGCCAG	CTATCAACTC	CGCCCTGGG	AGGGATTTTT
mxoxox3	CGCCATCTGC	CACCCAGCCAG	CTATCAACTC	CGCCCTGGG	AGGGATTTTT
mxoxox4	CGCCATCTGC	CACCCAGCCAG	CTATCAACTC	CGCCCTGGG	AGGGATTTTT
mxoxox5	CGCCATCTGC	CACCCAGCCAG	CTATCAACTC	CGCCCTGGG	AGGGATTTTT
mxoxox6	CGCCATCTGC	CACCCAGCCAG	CTATCAACTC	CGCCCTGGG	AGGGATTTTT
746				785	
wtCre	GAAGCAACTC	ATCGATTGAT	TTACGGCGCT	AAGGATGACT	CTGGTCAGAG
mxoxox1	GAAGCAACTC	ATCGATTGAT	TTACGGCGCT	AAGGATGACT	CTGGTCAGAG
mxoxox2	GAAGCAACTC	ATCGATTGAT	TTACGGCGCT	AAGGATGACT	CTGGTCAGAG
mxoxox3	GAAGCAACTC	ATCGATTGAT	TTACGGCGCT	AAGGATGACT	CTGGTCAGAG
mxoxox4	GAAGCAACTC	ATCGATTGAT	TTACGGCGCT	AAGGATGACT	CTGGTCAGAG
mxoxox5	GAAGCAACTC	ATCGATTGAT	TTACGGCGCT	AAGGATGACT	CTGGTCAGAG
mxoxox6	GAAGCAACTC	ATCGATTGAT	TTACGGCGCT	AAGGATGACT	CTGGTCAGAG
796					

Table 10 (cont)

*In Vitro* Evolution of the Cre recombinase

wtCre	ATACCTGGCC	TGGTCTGGAC	ACAGTGCOCG	TGTGGAGGCC	GCGCGAGATA
mx0x0x1	ATACCTGGCC	TGGTCTGGAC	ACAGTGCOCG	TGTGGAGGCC	GCGCGAGATA
mx0x0x2	ATACCTGGCC	TGGTCCGGAC	ACAGTGCOCG	TGTGGAGGCC	GCGCGAGATA
mx0x0x3	ATACCTGGCC	TGGTCCGGAC	ACAGTGCOCG	TGTGGAGGCC	GCGCGAGATA
mx0x0x4	ATACCAAGGCC	TGGTCTGGAC	ACAGTGCOCG	TGTGGAGGCC	GCGCGAGATA
mx0x0x5	ATACCTGGCC	TGGTCTGGAC	ACAGTGCOCG	TGTGGAGGCC	GCGCGAGATA
mx0x0x6	ATACCAAGGCC	TGGTCTGGAC	ACAGTGCOCG	TGTGGAGGCC	GCGCGAGATA
846					
wtCre	TGGCCCCGCC	TGGAGTTTCA	ATACCGGAGA	TCATGCAAGC	TGGTGGCTGG
mx0x0x1	TGGCCCCGCC	TGGAGTTTCA	ATACCGGAGA	TCATGCAAGC	TGGTGGCTGG
mx0x0x2	TGGCCCCGCC	TGGAGTTTCA	ATACCGGAGA	TCATGCAAGC	TGGTGGCTGG
mx0x0x3	TGGCCCCGCC	TGGAGTTTCA	ATACCGGAGA	TCATGCAAGC	TGGTGGCTGG
mx0x0x4	TGGCCCCGCC	TGGAGTTTCA	ATACCGGAGA	TCATGCAAGC	TGGTGGCTGG
mx0x0x5	TGGCCCCGCC	TGGAGTTTCA	ATACCGGAGA	TCATGCAAGC	TGGTGGCTGG
mx0x0x6	TGGCCCCGCC	TGGAGTTTCA	ATACCGGAGA	TCATGCAAGC	TGGTGGCTGG
896					
wtCre	ACCAATGTAA	ATATTGTTCAT	GAACATATC	CGTAACCTGG	ATAGTGA AAC
mx0x0x1	ACCAATGTAA	ATATTGTTCAT	GAACATATC	CGTAACCTGG	ATAGTGA AAC
mx0x0x2	TCCAATGTAA	ATATTGTTCAT	GAACATATC	CGTAACCTGG	ATAGTGA AAC
mx0x0x3	ACCAATGTAA	ATATTGTTCAT	GAACATATC	CGTAACCTGG	ATAGTGA AAC
mx0x0x4	ACCAATGTAA	ATATTGTTCAT	GAACATATC	CGTAACCTGG	ATAGTGA AAC
mx0x0x5	TCCAATGTAA	ATATTGTTCAT	GAACATATC	CGTAACCTGG	ATAGTGA AAC
mx0x0x6	TCCAATGTAA	ATATTGTTCAT	GAACATATC	CGTAACCTGG	ATAGTGA AAC
946					
wtCre	AGGGGCAATG	GTGGGCTGCG	TGGAAGATGG	CGATTAGCCA	TTAAGCGGTA
mx0x0x1	AGGGGCAATG	GTGGGCTGCG	TGGAAGATGG	CGATTAGCCA	TTAAGCGGTA
mx0x0x2	AGGGGCAATG	GTGGGCTGCG	TGGAAGATGG	CGATTAGCCA	TTAAGCGGTA
mx0x0x3	AGGGGCAATG	GTGGGCTGCG	TGGAAGATGG	CGATTAGCCA	TTAAGCGGTA
mx0x0x4	AGGGGCAATG	GTGGGCTGCG	TGGAAGATGG	CGATTAGCCA	TTAAGCGGTA
mx0x0x5	AGGGGCAATG	GTGGGCTGCG	TGGAAGATGG	CGATTAGCCA	TTAAGCGGTA
mx0x0x6	AGGGGCAATG	GTGGGCTGCG	TGGAAGATGG	CGATTAGCCA	TTAAGCGGTA
996					
wtCre	AATGATAAGC	TGGCTGTTT	TGGCGGATGA	GAGAAGATT	TCAGCTGAT
mx0x0x1	AATGATAAGC	TGGCTGTTT	TGGCGGATGA	GAGAAGATT	TCAGCTGAT
mx0x0x2	AATGATAAGC	TGGCTGTTT	TGGCGGATGA	GAGAAGATT	TCAGCTGAT
mx0x0x3	AATGATAAGC	TGGCTGTTT	TGGCGGATGA	GAGAAGATT	TCAGCTGAT
mx0x0x4	AATGATAAGC	TGGCTGTTT	TGGCGGATGA	GAGAAGATT	TCAGCTGAT
mx0x0x5	AATGATAAGC	TGGCTGTTT	TGGCGGATGA	GAGAAGATT	TCAGCTGAT
mx0x0x6	AATGATAAGC	TGGCTGTTT	TGGCGGATGA	GAGAAGATT	TCAGCTGAT
1046					
wtCre	ACAGATTAAA	TCAGAAAGCA	GA		
mx0x0x1	ACAGATTAAA	TCAGAAAGCA	GA		
mx0x0x2	ACAGATTAAA	TCAGAAAGCA	GA		
mx0x0x3	ACAGATTAAA	TCAGAAAGCA	GA		
mx0x0x4	ACAGATTAAA	TCAGAAAGCA	GA		
mx0x0x5	ACAGATTAAA	TCAGAAAGCA	GA		
mx0x0x6	ACAGATTAAA	TCAGAAAGCA	GA		
1096					

A T 2 - Start Codon  
T A 2 - Stop Codon  
T - Point Mutation

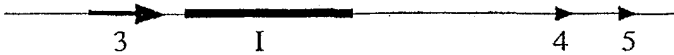
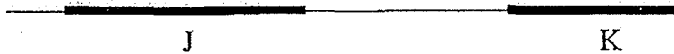
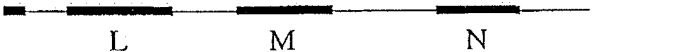
Table 11

*In Vitro* Evolution of the Cre recombinase

wtCre	MSNLLTVHQN	LPALPVDATS	DEVKKNLMDM	FRDRQAESBH	IWKMLLSVCB
mxoxox1	MSNLLTVHQN	LPALPVDATS	DEVKKNLMAM	FRDRQAESBH	IWKMLLSVCB
mxoxox2	MSNLLTVHQN	LPALPVDATS	DEVKKNLMDM	FRDRQAESBH	IWKMLLSVCB
mxoxox3	MSNLLTVHQN	LPALPIDATS	DEVKKNLMDM	FRDRQAESBH	IWKMLLSVCB
mxoxox4	MSNLLTVHQN	LPALPVDATS	DEVKKNLMDM	FRDRQAESBH	IWKMLLSVCB
mxoxox5	MSNLLTVHQN	LPALPVDATS	DEVKKNLMAM	FRDRQAESBH	IWKMLLSVCB
mxoxox6	MSNLLTVHQN	LPALPVDATS	DEVKKNLMDM	FRDRQAESBH	IWKMLLSVCB
1					
A					
wtCre	SWAAWCKLNN	RKWFPAPED	VRDYLLYLQA	BQLAVKTIQQ	HLGQLNMLHB
mxoxox1	SWAAWCKLNN	RKWFPAPED	VRDYLLYLQA	BQLAVKTIQQ	HLGQLNMLHB
mxoxox2	SWAAWCKLNN	RKWFPAPED	VRDYLLYLQA	BQLAVKTIQQ	HLGQLNMLHB
mxoxox3	SWAAWCKLNN	RKWFPAPED	VRDYLLYLQA	BQLAVKTIQQ	HLGQLNMLHB
mxoxox4	SWAAWCKLNN	RKWFPAPED	VRDYLLYLQA	BQLAVKTIQQ	HLGQLNMLHB
mxoxox5	SWAAWCKLNN	RKWFPAPED	VRDYLLYLQA	BQLAVKTIQQ	HLGQLNMLHB
mxoxox6	SWAAWCKLNN	RKWFPAPED	VRDYLLYLQA	BQLAVKTIQQ	HLGQLNMLHB
51					
C					
wtCre	BSGLPRPSDS	NAVSLVMBRI	BKENVDAGER	AKQALAFERT	DFDQVRSIME
mxoxox1	BSGLPRPSDS	NAVSLVMBRI	BKENVDAGER	AKQALAFERT	DFDQVRSIME
mxoxox2	BSGLPRPSDS	NAVSLVMBRI	BKENVDAGER	AKQALAFERT	DFDQVRSIME
mxoxox3	BSGLPRPSDS	NAVSLVMBRI	BKENVDAGER	AKQALAFERT	DFDQVRSIME
mxoxox4	BSGLPRPSDS	NAVSLVMBRI	BKENVDAGER	AKQALAFERT	DFDQVRSIME
mxoxox5	BSGLPRPSDS	NAVSLVMBRI	BKENVDAGER	AKQALAFERT	DFDQVRSIME
mxoxox6	BSGLPRPSDS	NAVSLVMBRI	BKENVDAGER	AKQALAFERT	DFDQVRSIME
101					
E					
wtCre	NSDBCCQDIEN	LAFLGIAYNT	LLFAEIAI	RVKDISRTDG	GRMLIHIGRT
mxoxox1	NSDBCCQDIEN	LAFLGIAYNT	LLFAEIAI	RVKDISRTDG	GRMLIHIGRT
mxoxox2	NSDBCCQDIEN	LAFLGIAYNT	LLFAEIAI	RVKDISRTDG	GRMLIHIGRT
mxoxox3	NSDBCCQDIEN	LAFLGIAYNT	LLFAEIAI	RVKDISRTDG	GRMLIHIGRT
mxoxox4	NSDBCCQDIEN	LAFLGIAYNT	LLFAEIAI	RVKDISRTDG	GRMLIHIGRT
mxoxox5	NSDBCCQDIEN	LAFLGIAYNT	LLFAEIAI	RVKDISRTDG	GRMLIHIGRT
mxoxox6	NSDBCCQDIEN	LAFLGIAYNT	LLFAEIAI	RVKDISRTDG	GRMLIHIGRT
151					
G					
H					
I					
2					

Table II (cont)

*In Vitro* Evolution of the Cre recombinase

wtCre	KTLVSTAGVE	KALSLGVTKL	VERWISVSGV	ADDPNNYLFC	RVRKNGVAAP
mx0x0x1	KTLVSTAGVE	KALSLGVTKL	VERWISISGV	ADDPNNYLFC	RVRKNGVAAP
mx0x0x2	KTLVSTAGVE	KALSLGVTKL	VERWISVSGV	ADDPNNYLFC	RVRKNGVAAP
mx0x0x3	KTLVSTAGVE	KALSLGVTKL	VEQWISVSGV	ADDPNNYLFC	RVRKNGVAAP
mx0x0x4	KTLVSTAGVE	KALSLGVTKL	VERWISVSGV	ADDPNNYLFC	RVRKNGVAAP
mx0x0x5	KTLVSTAGVE	KALSLGVTKQ	VERWISVSGV	ADDPNNYLFC	RVRKNGVAAP
mx0x0x6	KTLVSTAGVE	KALSLGVTKL	VERWISVSGV	ADDPNNYLFC	RVRKNGVAAP
201					
wtCre	SATSQLSBTA	LEGIFEATHR	LIYGAKDDSG	QBYLAWSGHS	AFVGAARDMA
mx0x0x1	SATSQLSBTA	LGGIFEATHR	LIYGAKDDSG	QBYLAWSGHS	AFVGAARDMA
mx0x0x2	SATSQLSBTA	LGGIFEATHR	LIYGAKDDSG	QBYLAWSGHS	AFVGAARDMA
mx0x0x3	SATSRLSBTA	LGGIFEATHR	LIYGAKDDSG	QBYLAWSGHS	AFVGAARDMA
mx0x0x4	SATSQLSBTA	LEGIFEATHR	LIYGAKDDSG	QBYQAWSGHS	AFVGAARDMA
mx0x0x5	SATSQLSBTA	LGGIFEATHR	LIYGAKDDSG	QBYLAWSGHS	AFVGAARDMA
mx0x0x6	SATSQLSBTA	LGGIFEATHR	LIYGAKDDSG	QBYQAWSGHS	AFVGAARDMA
251					
wtCre	RAGVSIPEIM	QAGGWITNVI	VMNVTBNLDS	ETGAMVRLLE	DGD*
mx0x0x1	RAGVSIPEIM	QAGGWITNVI	VMNVTBNLDS	ETGAMVRLLE	DGD*
mx0x0x2	RAGVSIPEIM	QAGGWSNVNI	VMNVTBNLDS	ETGAMVRLLE	DGD*
mx0x0x3	RAGVSIPEIM	QAGGWITNVI	VMNVTBNLDS	ETGAMVRLLE	DGD*
mx0x0x4	RAGVSIPEIM	QAGGWITNVI	VMNVTBNLDS	ETGAMVRLLE	DGD*
mx0x0x5	RAGVSIPEIM	QAGGWSNVNI	VMNVTBNLDS	ETGAMVRLLE	DGD*
mx0x0x6	RAGVSIPEIM	QAGGWSNVNI	VMNVTBNLDS	ETGAMVRLLE	DGD*
301					

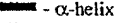
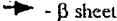
G - mutated residue  
 T - catalytic residue  
 E - DNA contacting residue  
 -  $\alpha$ -helix  
 -  $\beta$  sheet

Table 12

Transition	Frequency	Transversion	Frequency
A - G	29 % (9)	A - C	6.5 % (2)
G - A	26 % (8)	C - A	< 3 % (0)
C - T	6.5 % (2)	A - T	9.7 % (3)
T - C	6.5 % (2)	T - A	9.7 % (3)
		C - G	< 3 % (0)
		G - C	< 3 % (0)
		G - T	6.5 % (2)
		T - G	< 3 % (0)

31 point mutations could be identified in the six analyzed mutants. The mutagenic frequency can therefore be calculated as 0.5%. No frame-shift mutations, or more than one point mutation per codon of the *cre* sequence were found. The table classifies these mutations into the different types of transition and transversion events. Given are the frequencies and in parenthesis the actual numbers which were found. Interestingly, the A to G and vice versa transision occurred much more often than all other possible events. On the other hand, half of the possible transversions were not identified at all.

	Round	Induction	% Kan <sup>R</sup>	Nb. Kan <sup>R</sup>	% White
5	0	4 h	< 0.01%	-	100%
	1	4 h	0.16%	6	84%
10	2	4 h	0.56%	47	82%
	3	2.5 h	0.2%	36	80%
15	4	2.5 h	4.6%	102	75%

20

25

30

35

40

45

50

55

5

% Kan<sup>R</sup>

10

15

20

25

30

35

40

45

50

55

Mutant	<i>loxK2</i> <sup>2</sup>	<i>loxP</i> <sup>2</sup>
wt	<0.01%	94%
R3M1	66%	98%
R3M2	54%	98%
R3M3	65%	91%
R3M4	0.02%	93%
R3M5	15%	91%
R3M6	3%	94%

Table 15 In Vitro Recombination Results

% of recombination is calculated from the obtained fluorescence intensity of substrate and product on an agarose gel.

Cre	% of lox-lox recombination <i>In vitro</i>											
	loxP			ATAACTTCGTATA								
	TT	TG	GT	GG	TC	CC	AA	loxK2	loxK1	loxP-K2	loxP-K1	
R3M3	25.4	32.1	36.8	29.7	8.5	23.9	36.7	29.8	31.2	28.9		
E262G/D29A	25.4	30.4	21.4	48.4	16.6	20.4	52.4	17.7	31.0	29.7		
E262G/T316S	36.4	37.6	38.6	47.3	19.6	23.2	51.1	32.3	43.2	29.1		
E262G/D189N	35.3	33.8	34.0	59.9	16.4	24.0	65.1	31.2	46.5	35.3		
E262G	30.9	32.1	27.5	49.2	-	15.2	46.9	24.3	35.9	28.4		
wt	28.7	0	45.8	0	0	0	0	0	15.9	4.3		

Table 16 DNA Binding to lox halfsites (Gelshift)

% of gel retardation is calculated from the obtained intensity of the shifted band (Cre bound to lox halfsite) and free lox halfsites (labeled oligonucleotides). The values given here were all obtained with 0.2 pmol of lox halfsites. Without enzyme added, no retardation of either halfsite could be observed. The obtained data was normalized for better comparability to 100% retardation on loxP halfsite, for each enzyme tested.

% of lox halfsites retarded on the gel due to Cre binding

	loxP	loxK2	loxK1
R3M3	100	14.8 ± 5.2	0.3 ± 0.2
E262G/D29A	100	36.6 ± 6.4	14.1 ± 2.7
E262G/T316S	100	40.7 ± 19.3	13.6 ± 1.7
E262G/D189N	100	44.2 ± 8.1	16.1 ± 2.9
E262G	100	34 ± 7.2	15.8 ± 6.4
wt	100	6 ± 1.3	0.8 ± 1.1
	ATAACTTCGTATA	ATAACAACGTATA	ATACCTTTGTATA

Table 17

Name	pBS#	Sequence: Halfsite - FAS1 Spacer - Inverted Halfsite									
Positions in 1ox:		17	12	5	4	1/1'4'	5'	11'	17'		
1oxP	632	ATAACTTCGTATA TACC/TTTC TATAGCAAGTTAT									
1oxK2	633	ATPACAAAGTATA TACC/TTTC TATAGCTTGTAT									
1oxK1	634	ATACCTTTGTATA TACC/TTTC TATAGAAAGGTAT									
1oxK2 'GG'	638	ATAACGGCGTATA TACC/TTTC TATAGCCCCTTAT									
1oxK2 'CC'	639	ATACCCCGTATA TACC/TTTC TATAGCGGGTTAT									
1oxK2 'TC'	635	ATAACTCCGTATA TACC/TTTC TATAGCGAGTTAT									
1oxK2 'GT'	636	ATAACGTCGTATA TACC/TTTC TATAGCACGTTAT									
1oxK2 'TG'	637	ATACACGCGTATA TACC/TTTC TATAGCCAGTTAT									

## Claims

5

10

15

20

25

30

35

40

45

50

55

## CLAIMS

5

We claim:

10

1. A method of identifying variant recombinases that mediate recombination at variant recombination sites, the method comprising,

15

(a) bringing into contact

a mutant recombinase,

a first nucleic acid sequence comprising a first reporter gene and first and second recombination sites, wherein the first and second recombination sites are variant recombination sites, and

20

a second nucleic acid sequence comprising a second reporter gene and third and fourth recombination sites, wherein the third and fourth recombination sites can be recombined by a non-mutant recombinase,

25

(b) determining if recombination occurs between the first and second recombination sites, and determining if recombination occurs between the third and fourth recombination sites,

30

wherein recombination between the first and second recombination sites indicates that the mutant recombinase is a variant recombinase that mediates recombination at variant recombination sites,

wherein recombination between the third and fourth recombination sites indicates that the mutant recombinase retains the ability to mediate recombination at non-variant recombination sites.

35

2. The method of claim 1 wherein the recombination sites comprise recognition sequences and compatibility sequences,

40

wherein the recognition sequences of the first and second recombination sites differ from the recognition sequences of the third and fourth recombination sites,

wherein the compatibility sequences of the first and second recombination sites are sufficiently similar to allow recombination between the first and second recombination sites, and wherein the compatibility sequences of the third and fourth recombination sites are sufficiently similar to allow recombination between the third and fourth recombination sites, and

45

wherein the compatibility sequences of the first and second recombination sites differ from the compatibility sequences of the third and fourth recombination

50

55

5 sites such that neither the first nor the second recombination site can be recombined with either the third or the fourth recombination site.

3. The method of claim 1 wherein the first and second recombination sites cannot be recombined by non-mutant recombinase to a significant extent.

10 4. The method of claim 1 or 2 wherein the first and second recombination sites have identical sequences, and wherein the third and fourth recombination sites have identical sequences.

15 5. The method of claim 1 wherein recombination between the first and second recombination sites alters the expression of the first reporter gene, wherein recombination between the first and second recombination sites is determined by determining if expression of the first reporter gene is altered, and

20 wherein recombination between the third and fourth recombination sites alters the expression of the second reporter gene, wherein recombination between the third and fourth recombination sites is determined by determining if expression of the second reporter gene is altered.

25 6. The method of claim 5 wherein recombination between the first and second recombination sites allows the first reporter gene to be expressed.

30 7. The method of claim 6 wherein the first nucleic acid sequence further comprises a spacer sequence flanked by the first and second recombination sites, wherein the spacer sequence interrupts the first reporter gene such that the first reporter gene is not expressed, wherein recombination of the first and second recombination sites excises the spacer sequence which allows the first reporter gene to be expressed.

35 8. The method of claim 6 wherein a portion of the first reporter gene is inverted, wherein the inverted portion of the first reporter gene is flanked by the first and second recombination sites, wherein recombination of the first and second recombination sites inverts the inverted portion of the first reporter gene which allows the first reporter gene to be expressed.

40 9. The method of claim 5 wherein recombination between the first and second recombination sites prevents expression of the first reporter gene.

45 10. The method of claim 9 wherein the first reporter gene is flanked by the first and second recombination sites, wherein recombination of the first and second

5 recombination sites excises the first reporter gene which prevents expression of the first reporter gene.

11. The method of claim 9 wherein a portion of the first reporter gene is flanked by the first and second recombination sites, wherein recombination of the first and second recombination sites inverts the flanked portion of the first reporter gene which prevents expression of the first reporter gene.

12. The method of claim 5 wherein recombination between the third and fourth recombination sites allows the second reporter gene to be expressed.

13. The method of claim 12 wherein the second nucleic acid sequence further comprises a spacer sequence flanked by the third and fourth recombination sites, wherein the spacer sequence interrupts the second reporter gene such that the second reporter gene is not expressed, wherein recombination of the third and fourth recombination sites excises the spacer sequence which allows the second reporter gene to be expressed.

14. The method of claim 13 wherein the spacer sequence interrupts the second reporter gene such that the second reporter gene is not transcribed.

15. The method of claim 13 wherein the second reporter gene encodes a protein, wherein the spacer sequence interrupts the second reporter gene such that the protein encoded by the second reporter gene is not translated.

16. The method of claim 13 wherein the spacer sequence interrupts the second reporter gene such that the second reporter gene produces an inactive expression product.

17. The method of claim 12 wherein a portion of the second reporter gene is inverted, wherein the inverted portion of the second reporter gene is flanked by the third and fourth recombination sites, wherein recombination of the third and fourth recombination sites inverts the inverted portion of the second reporter gene which allows the second reporter gene to be expressed.

18. The method of claim 5 wherein recombination between the third and fourth recombination sites prevents expression of the second reporter gene to be expressed.

19. The method of claim 18 wherein the second reporter gene is flanked by the third and fourth recombination sites, wherein recombination of the third and

fourth recombination sites excises the second reporter gene which prevents expression of the second reporter gene.

20. The method of claim 18 wherein a portion of the second reporter gene is flanked by the third and fourth recombination sites, wherein recombination of the third and fourth recombination sites inverts the flanked portion of the second reporter gene which prevents expression of the second reporter gene.

21. The method of claim 1 wherein the first nucleic acid sequence is a first nucleic acid construct and the second nucleic acid sequence is on a second nucleic acid construct.

22. The method of claim 21 wherein the first nucleic acid construct is an extrachromosomal vector and the second nucleic acid construct is in the genome of a host cell.

23. The method of claim 1 wherein the first and second nucleic acid constructs are on the same nucleic acid construct.

24. A method for producing site-specific recombination of DNA, comprising,

contacting a variant recombinase identified by the method of claim 1 with first and second DNA sequences,

wherein the first DNA sequence comprises a first recombination site and the second DNA sequence comprises a second recombination site,

wherein the variant recombinase mediates recombination between the first and second recombination sites thereby producing the site specific recombination.

25. The method of claim 24 wherein the first recombination site, the second recombination site, or both, are variant recombination sites.

26. The method of claim 24, wherein the first and second DNA sequences are connected by a pre-selected DNA segment.

27. The method of claim 26, wherein the first and second recombination sites have the same orientation and the site-specific recombination of DNA is a deletion of the pre-selected DNA segment.

28. The method of claim 27, wherein the pre-selected DNA segment is a gene for a structural protein, an enzyme, or a regulatory molecule.

29. The method of claim 27 further comprising contacting the variant recombinase with a fourth DNA sequence comprising a third recombination site,

5 wherein the second and fourth DNA sequences are connected by a second pre-selected DNA segment.

10 30. The method of claim 29 wherein the first recombination site is a variant recombination site recognized by the variant recombinase and not by wild type recombinase, and wherein the second and third recombination sites are recombination sites recognized by wild type recombinase and by the variant recombinase.

15 31. The method of claim 30 further comprising, prior to contacting the variant recombinase with the first, second, and third recombination sites, contacting the recombination sites with wild type recombinase, thereby producing site specific recombination between the second and third recombination sites resulting in a deletion of the second pre-selected DNA segment.

20 32. The method of claim 29, wherein the second pre-selected DNA segment is a gene for a structural protein, an enzyme, or a regulatory molecule.

25 33. The method of claim 26, wherein the first and second recombination sites have opposite orientations and the site-specific recombination is an inversion of the nucleotide sequence of the pre-selected DNA segment.

30 34. The method of claim 33, wherein the first and second recombination sites are variant recombination sites recognized by the variant recombinase.

35 35. The method of claim 33, wherein the pre-selected DNA segment is a gene for a structural protein, an enzyme, or a regulatory molecule.

36. The method of claim 24, wherein the second and third DNA sequences are introduced into two different DNA molecules and the site-specific recombination is a reciprocal exchange of DNA segments proximate to the recombination sites.

40 37. The method of claim 36, wherein the first and second recombination sites are variant recombination sites recognized by the variant recombinase.

38. The method of claim 24 wherein the second DNA sequence includes a label, wherein recombination between the first and second recombination sites associates the label with the first DNA sequence.

45 39. The method of claim 38 wherein the first DNA sequence is a large circular DNA molecule.

50 40. The method of claim 24 wherein recombination occurs in a cell.

5 41. The method of claim 40 wherein the variant recombinase is contacted with the first and second DNA sequences by introducing into the cell a third DNA sequence comprising DNA encoding the variant recombinase.

10 42. The method of claim 41, wherein the third DNA sequence further comprises a regulatory nucleotide sequence and expression of the variant recombinase is produced by activating the regulatory nucleotide sequence.

15 43. The method of claim 40, wherein the cell is a eukaryotic cell, a mammalian cell, a yeast cell, a fungal cell, a prokaryotic cell, a bacterial cell, an archae bacterial cell, or a cell in a multicellular organism.

44. The method of claim 43 wherein the multicellular organism is a plant, an animal, or a mammal.

20 45. The method of claim 40, wherein the first and second DNA sequences are connected by a pre-selected DNA segment, wherein the first and second recombination sites have the same orientation and the site-specific recombination of DNA is a deletion of the pre-selected DNA segment.

25 46. The method of claim 45 wherein the cell is in a multicellular organism.

47. The method of claim 45, wherein the pre-selected segment is an undesired marker or trait gene.

30 48. The method of claim 24 wherein the variant recombinase is contacted with the recombination sites *in vitro*.

49. The method of claim 48 wherein the method further comprises introducing the recombined DNA into a cell.

35 50. A method for cloning large DNA fragments, the method comprising concatenating DNA fragments to be cloned with vector arms, wherein each vector arm comprises a recombination site, wherein the DNA fragments and vector arms alternate in the concatemer,

40 introducing the concatemer into a cell expressing a variant recombinase identified by the method of claim 1, wherein the recombinase mediates recombination of the recombination sites thereby generating circles each containing a DNA fragment and a vector arm.

45 51. A method for cloning large DNA fragments, the method comprising ligating a DNA fragment to be cloned to vector arms, wherein each vector arm comprises (i) a blunt end, (ii) another end which is compatible with an end of

the DNA fragment to be cloned, and (iii) a recombination site, wherein concatemers are not formed, and

introducing the ligated DNA fragment and vector arms into a cell expressing a variant recombinase identified by the method of claim 1.

52. A method for cloning large DNA fragments, the method comprising ligating a plurality of DNA fragments to be cloned with a plurality of first and second vector arms, wherein each first vector arm comprises two ligatable ends, wherein each second vector arm comprises a recombination site and one non-ligatable end,

wherein, following ligation, the DNA fragments and first vector arms alternate in concatemers, wherein the concatemers are flanked by second vector arms,

introducing the concatemers into a cell expressing a variant recombinase identified by the method of claim 1, wherein the recombinase mediates recombination of the recombination sites thereby generating circles containing the DNA fragments.

53. A variant recombinase identified by the method of claim 1.

54. A nucleic acid molecule encoding a variant recombinase identified by the method of claim 1.

55. The nucleic acid molecule of claim 54 wherein the nucleic acid molecule is a plasmid.

56. A cell containing the nucleic acid molecule of claim 54.

57. The cell of claim 56 wherein the cell is a eukaryotic cell, a mammalian cell, a yeast cell, a fungal cell, a prokaryotic cell, a bacterial cell, an archae bacterial cell, or a cell in a multicellular organism.

58. The cell of claim 57 wherein the multicellular organism is a plant, an animal, or a mammal.

59. A nucleic acid molecule having at least one variant recombination site, wherein the variant recombination site is recognized by a variant recombinase identified by the method of claim 1 and is not recognized by wild type recombinase.

60. The nucleic acid molecule of claim 59 wherein the recombination site is not recognized by wild type recombinase.

- 5           61. The nucleic acid molecule of claim 59, wherein a first recombination site  
and a second recombination site are connected by a pre-selected DNA segment.
62. The nucleic acid molecule of claim 59 wherein the nucleic acid molecule  
is a plasmid.
- 10           63. A cell containing the nucleic acid molecule of claim 59.

15

20

25

30

35

40

45

50

55

## loxP vs. loxK1 vs. loxK2

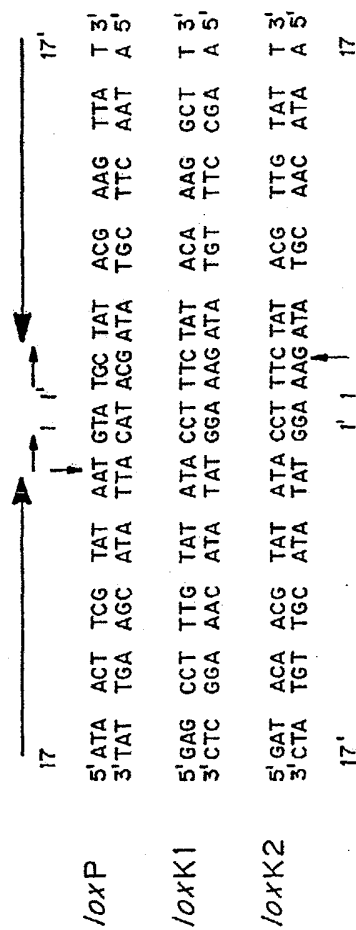


FIG. 1

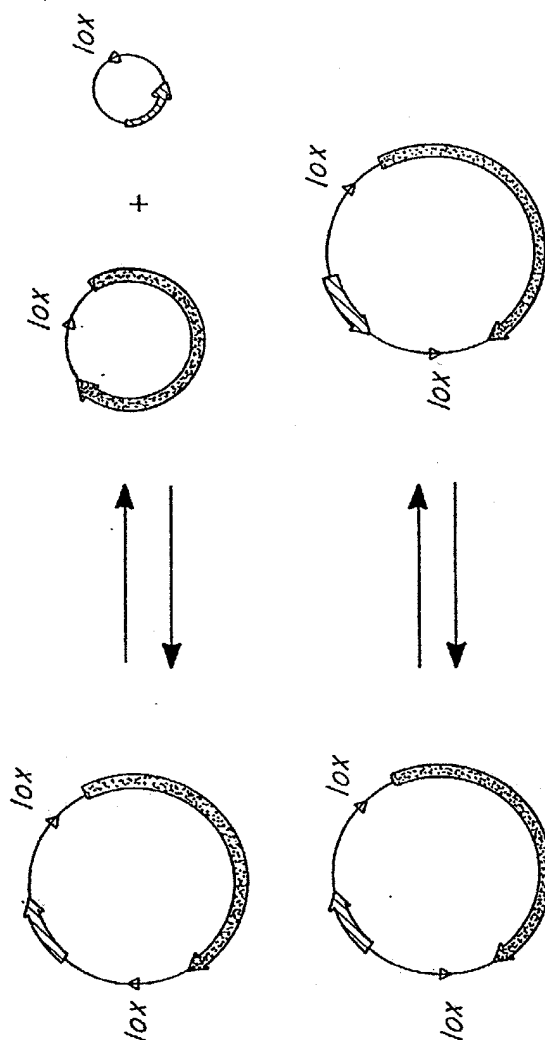


FIG. 2

3/25

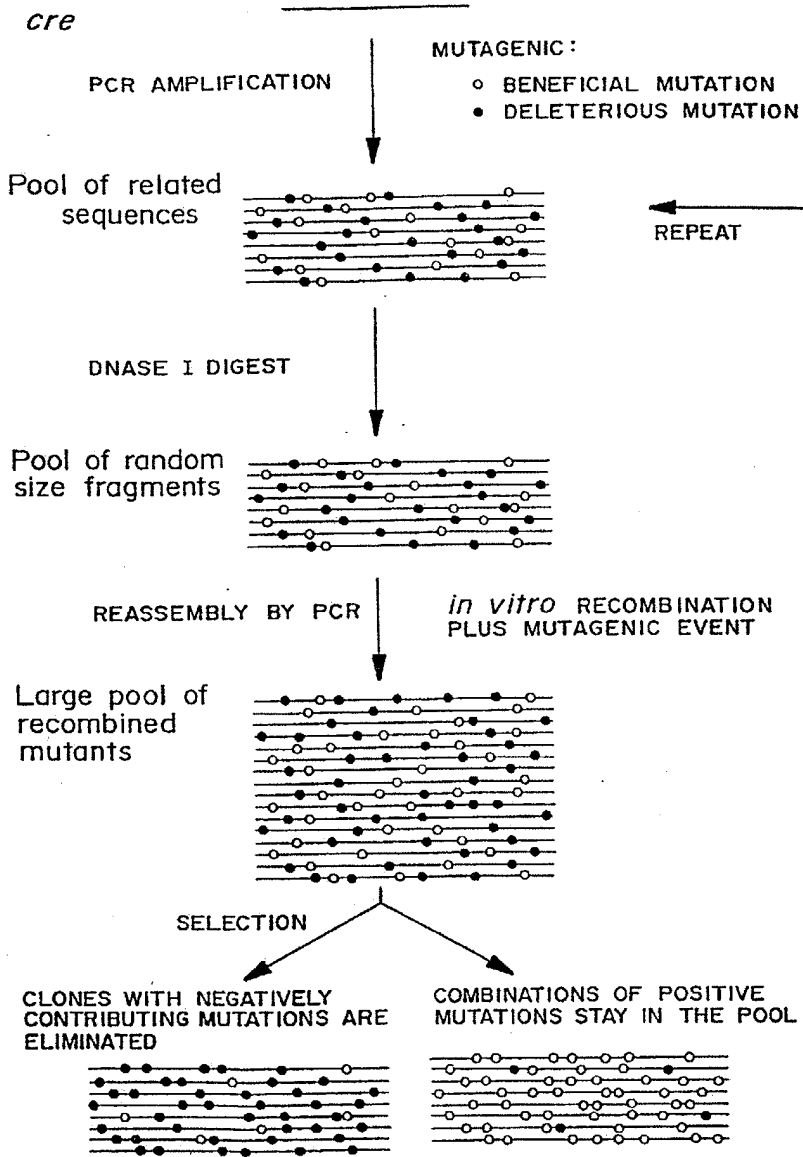
**FIG. 3**

FIG. 4

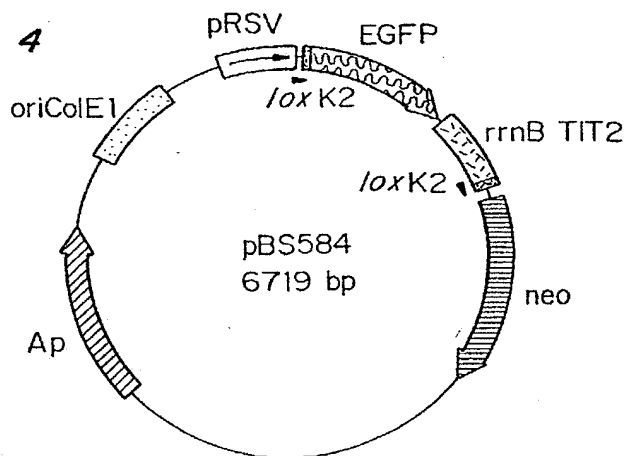
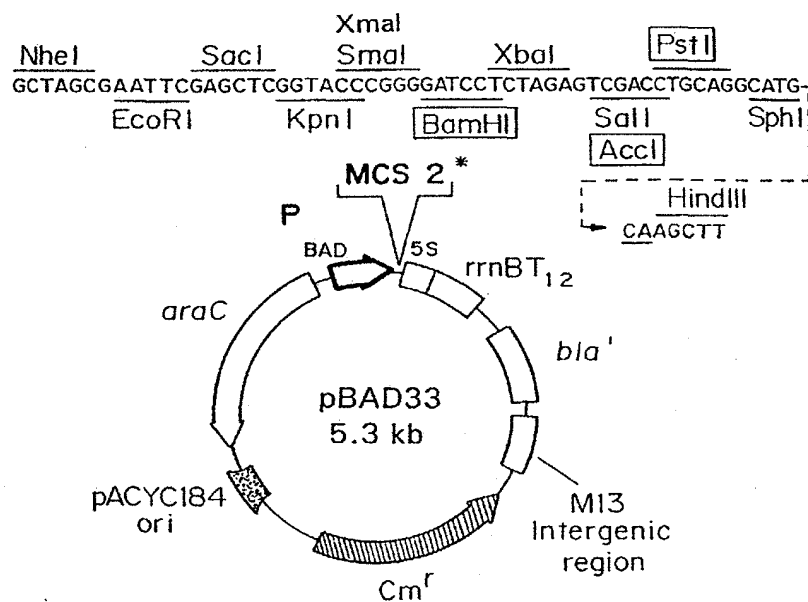


FIG. 6



**FIG. 5A**

100bp



**FIG. 5B**

300bp



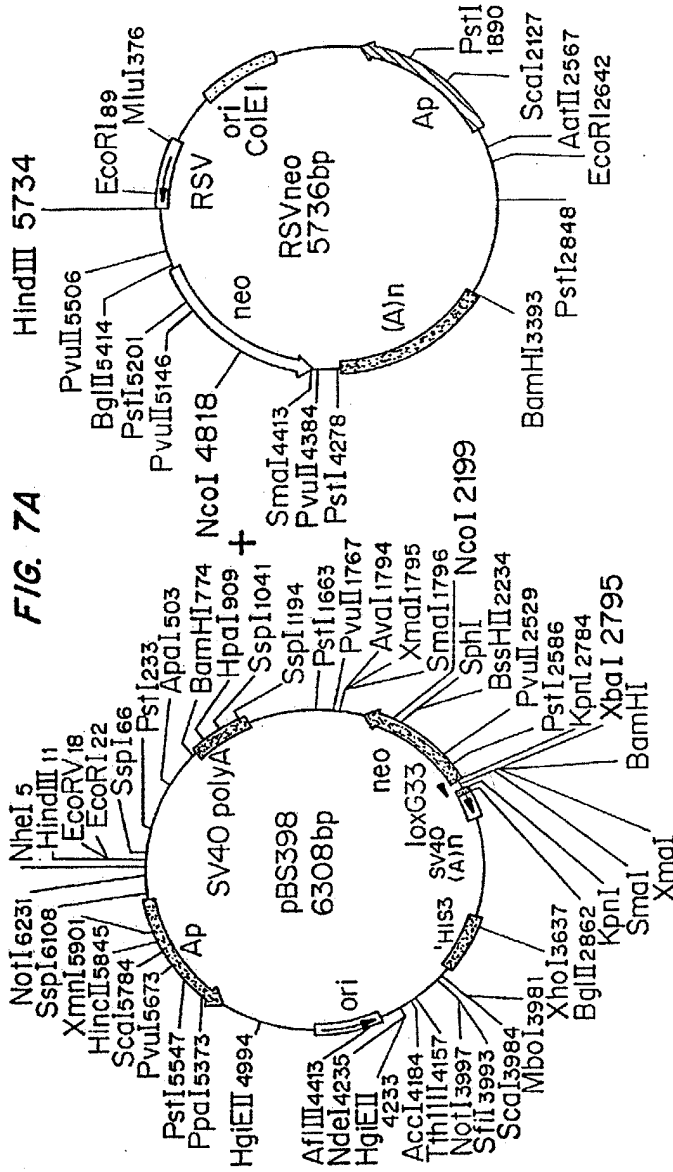
**FIG. 5C**

ca.2kb

300 bp

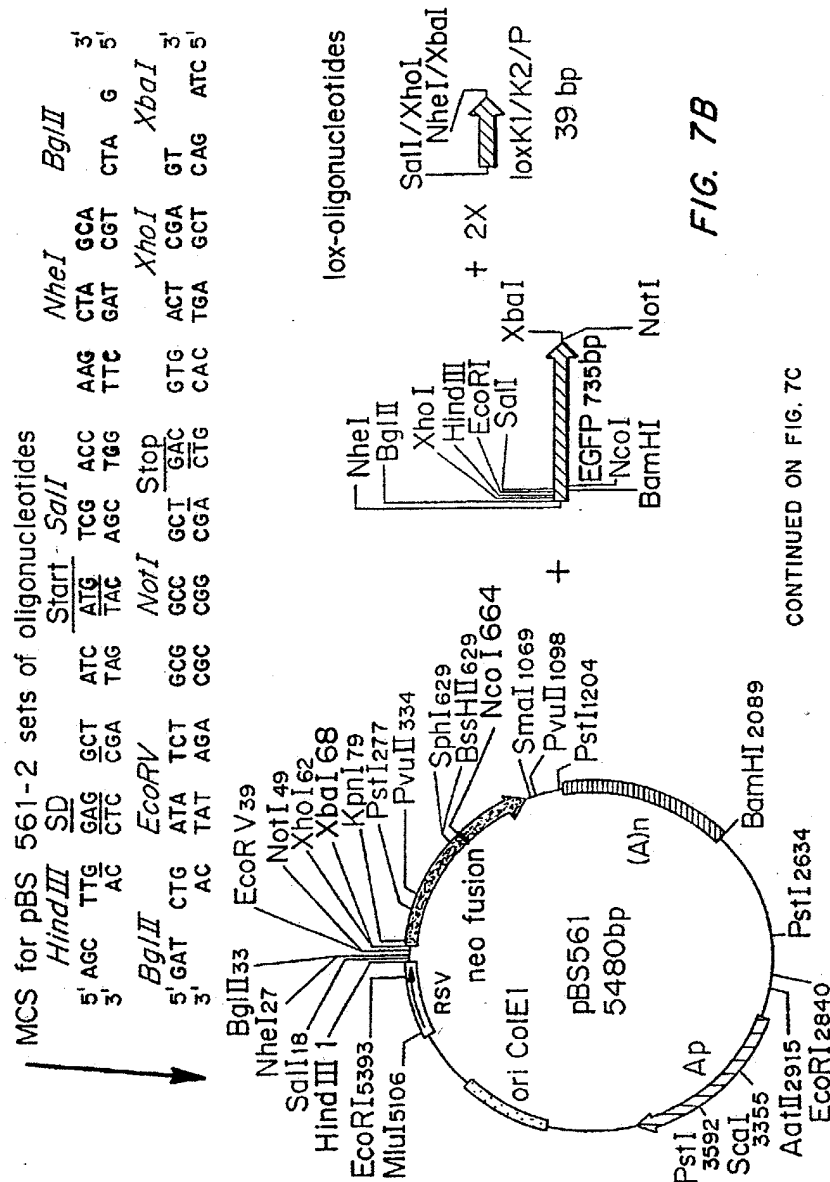


1.1kb cre



CONTINUED ON FIG. 7B

7/25



CONTINUED ON FIG. 7C

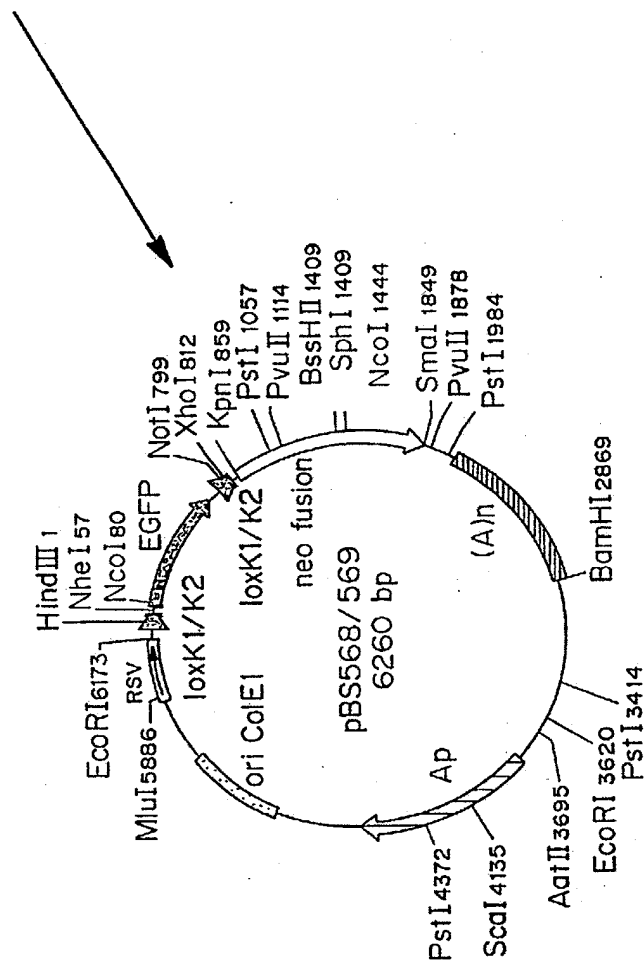
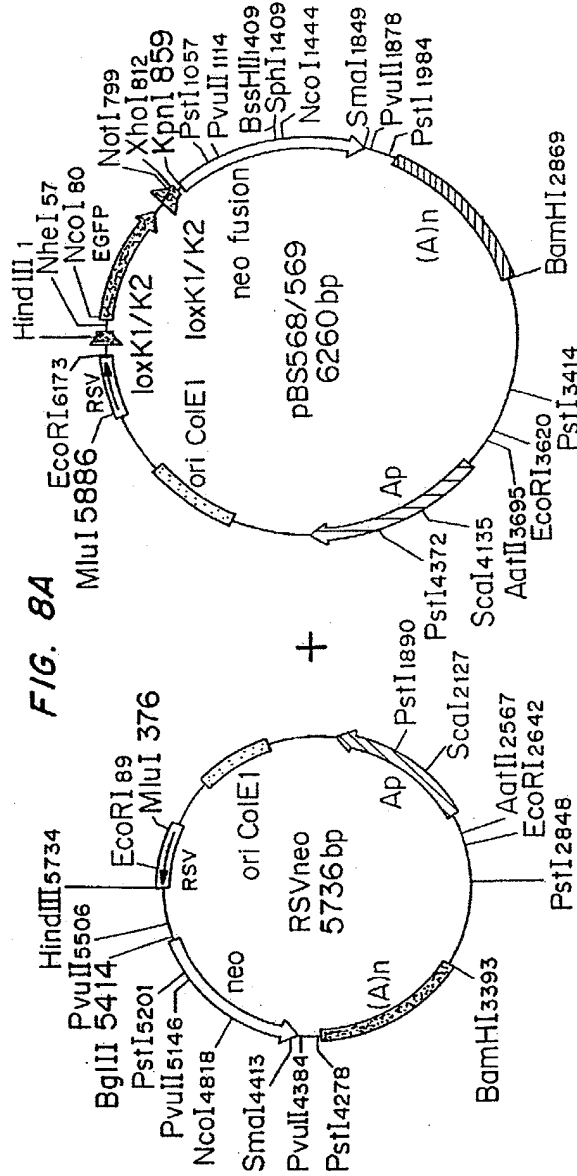


FIG. 7C

9/25

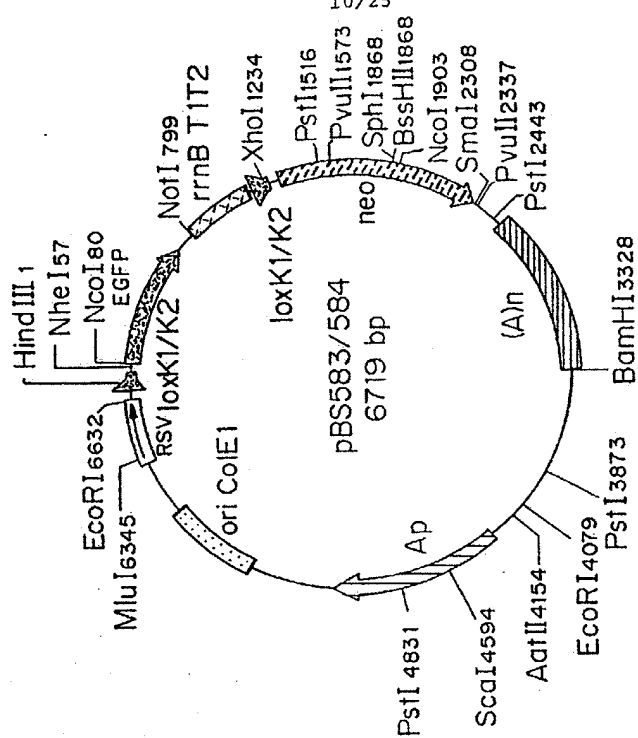


CONTINUED ON FIG. 8B

NotI 1 XhoI 436

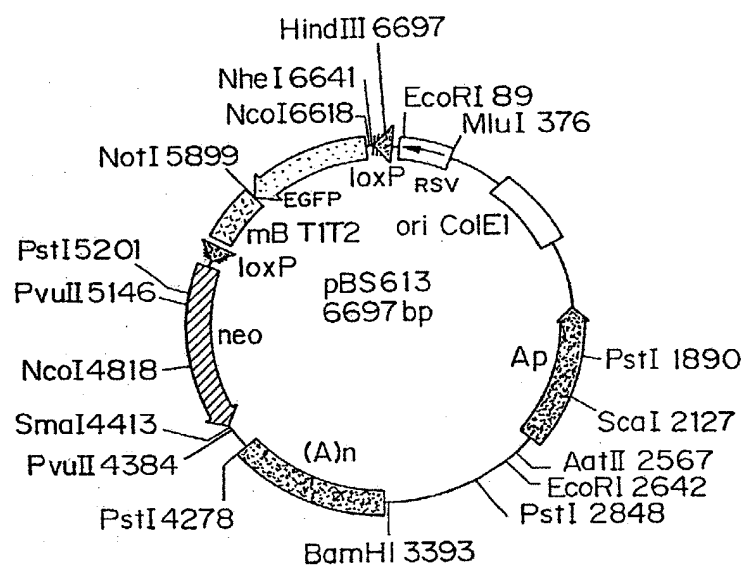
rrnB TIT2

10/25



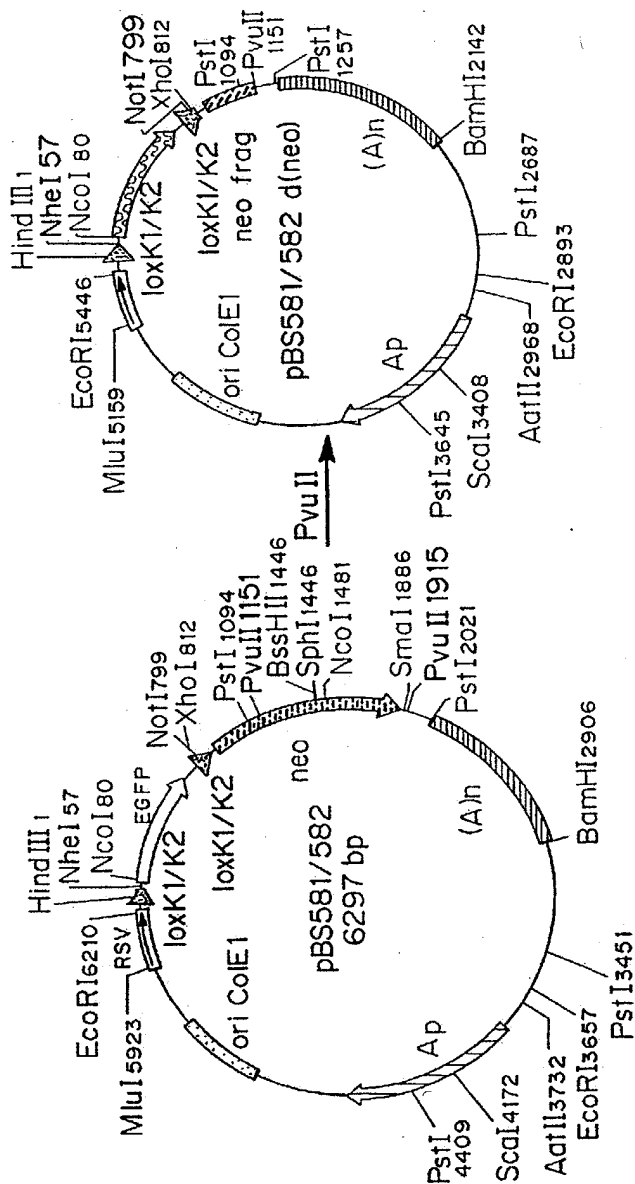
**FIG. 8B**

FIG. 9

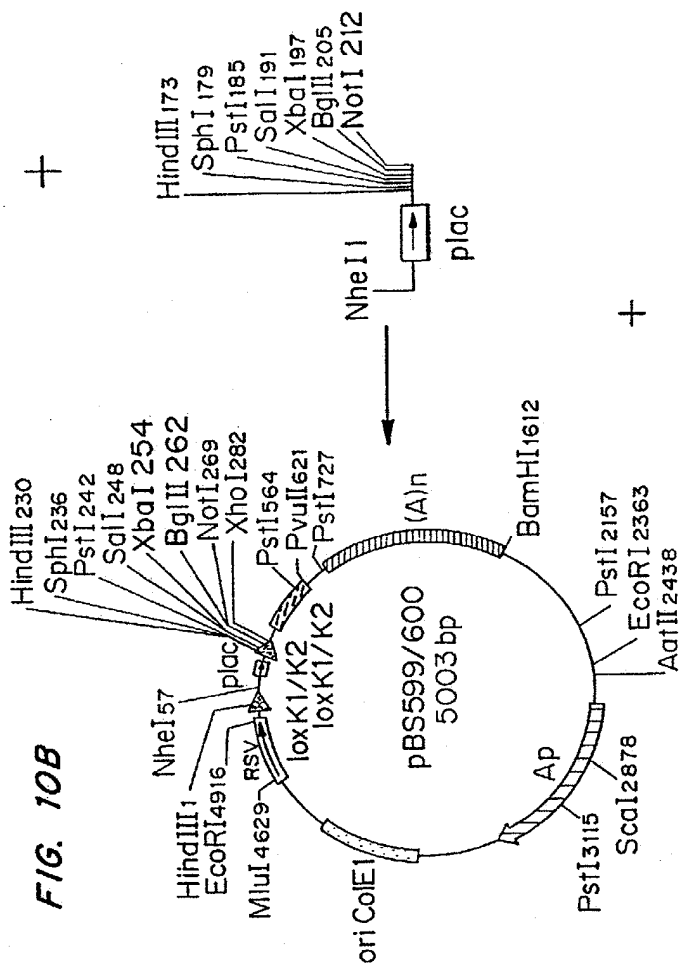


12/25

FIG. 10A



CONTINUED ON FIG. 10B



CONTINUED ON FIG. 10C

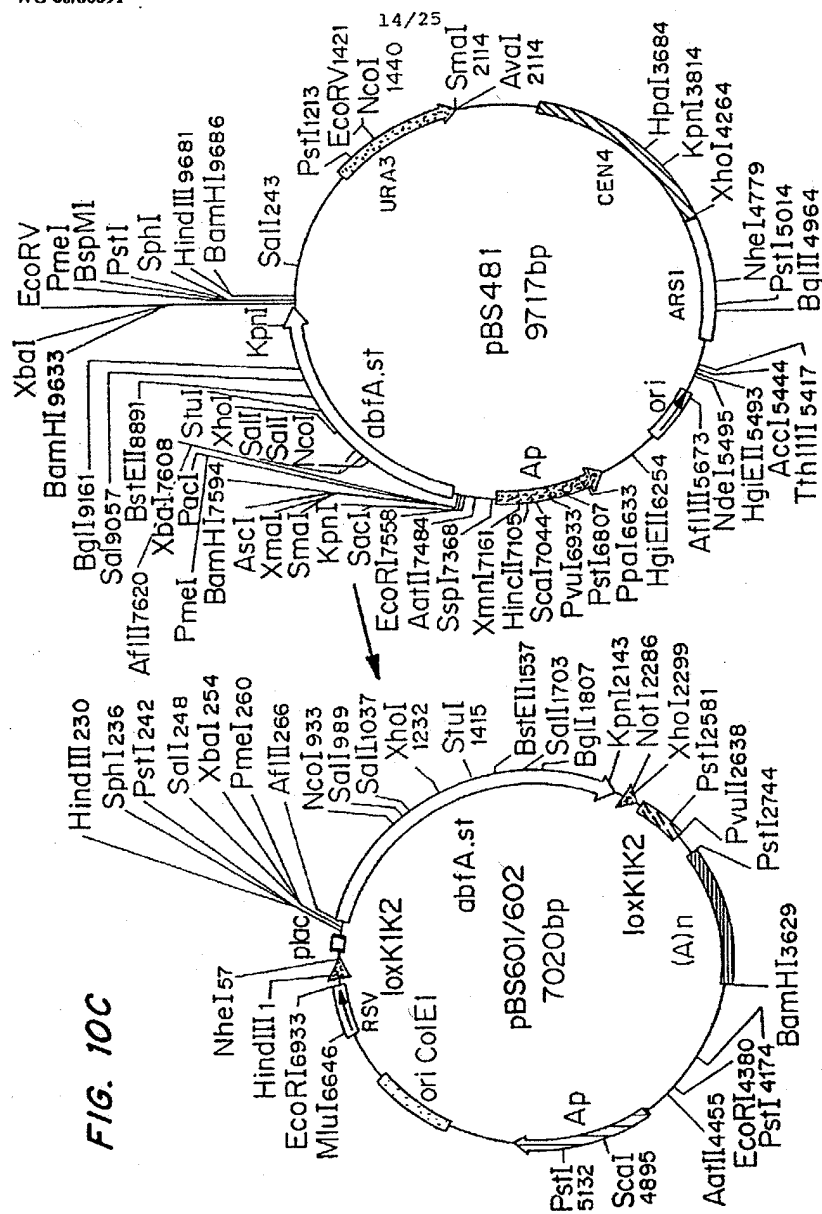
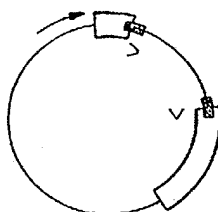
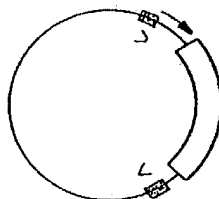


FIG. 11

INTERRUPTED  
(DELETION)



FLANKED  
(DELETION)



INVERTED  
(INVERSION)

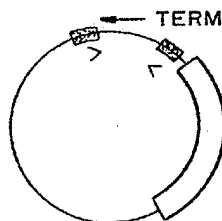
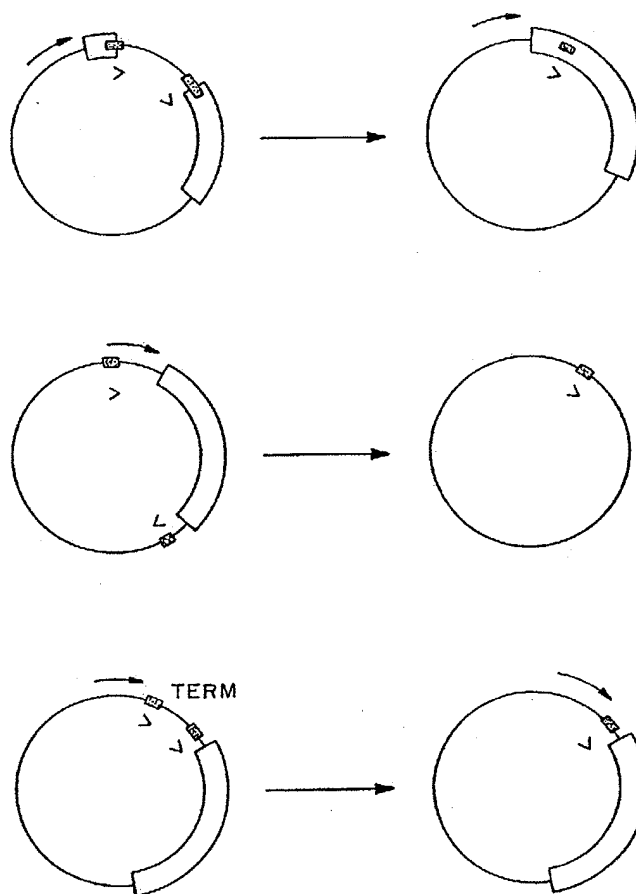
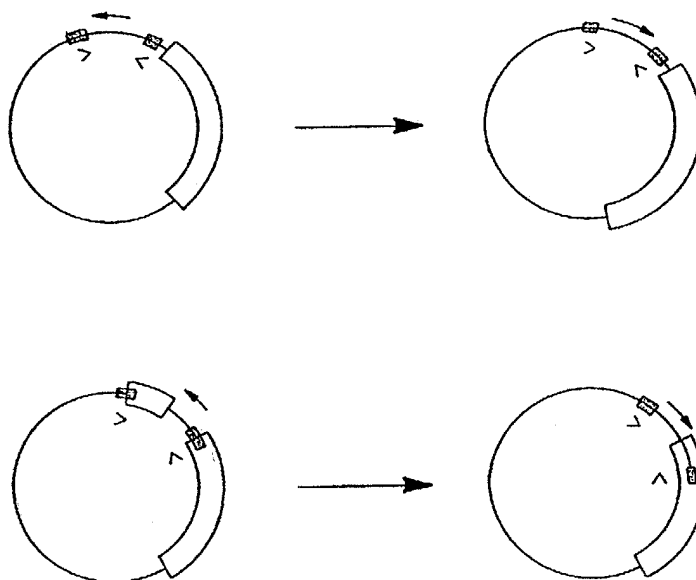


FIG. 12A



DELETION

FIG. 12B



INVERSION

FIG. 12C

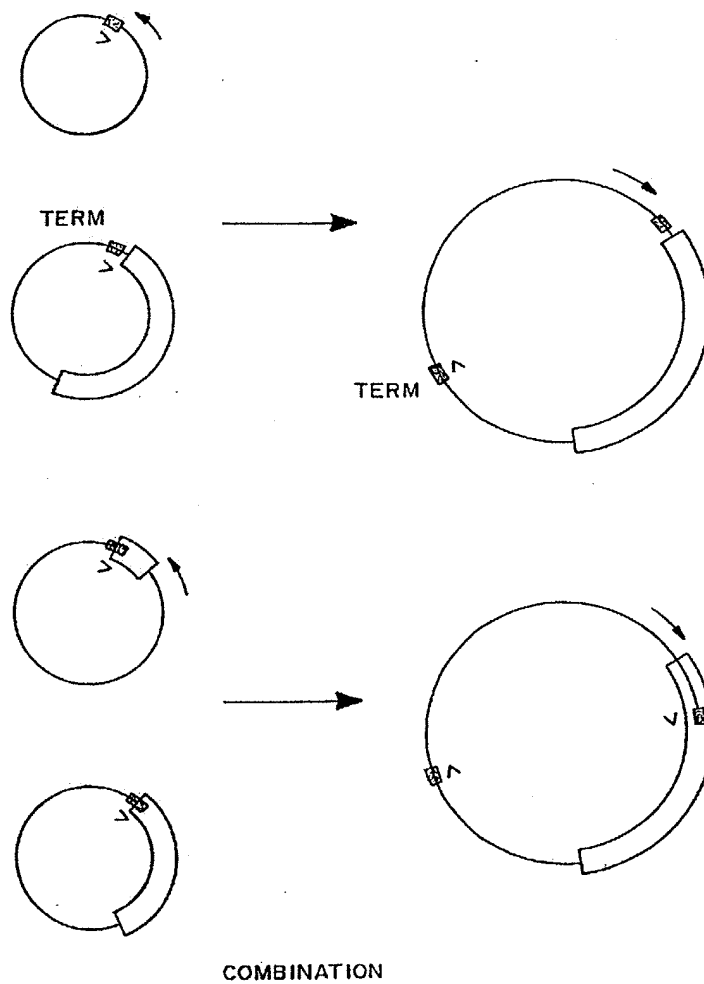
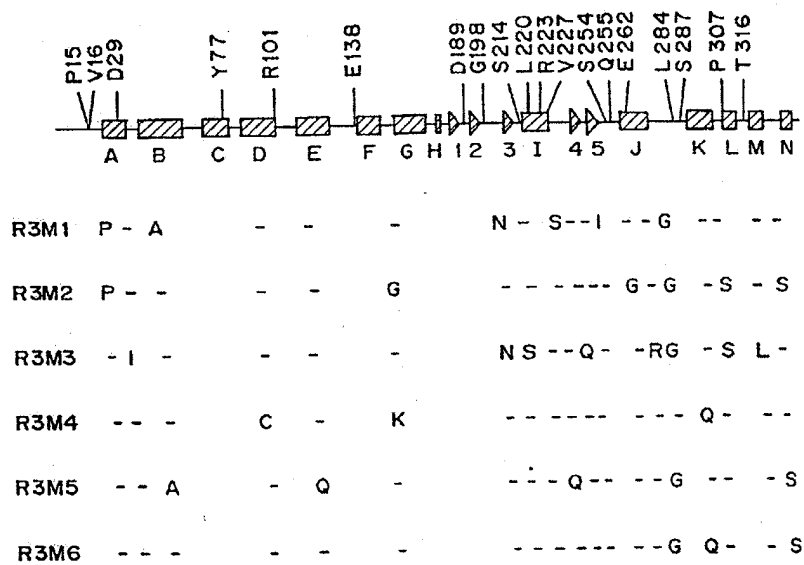


FIG. 13



**FIG. 14A**

% of Cre-mediated recombination

R3M3 Cre	97	82	67	90
E262G/D29A Cre	95	37	34	47
E262G/T316S Cre	98	21	14	32
E262G/D189N Cre	93	6	5	22
E262G Cre	90	20	10	28
wt Cre	95	0	0	0
	TT	GG	CC	AA

loxHalfsite      ATA A C T T C G T A T A

**FIG. 14B**

% of Cre-mediated recombination

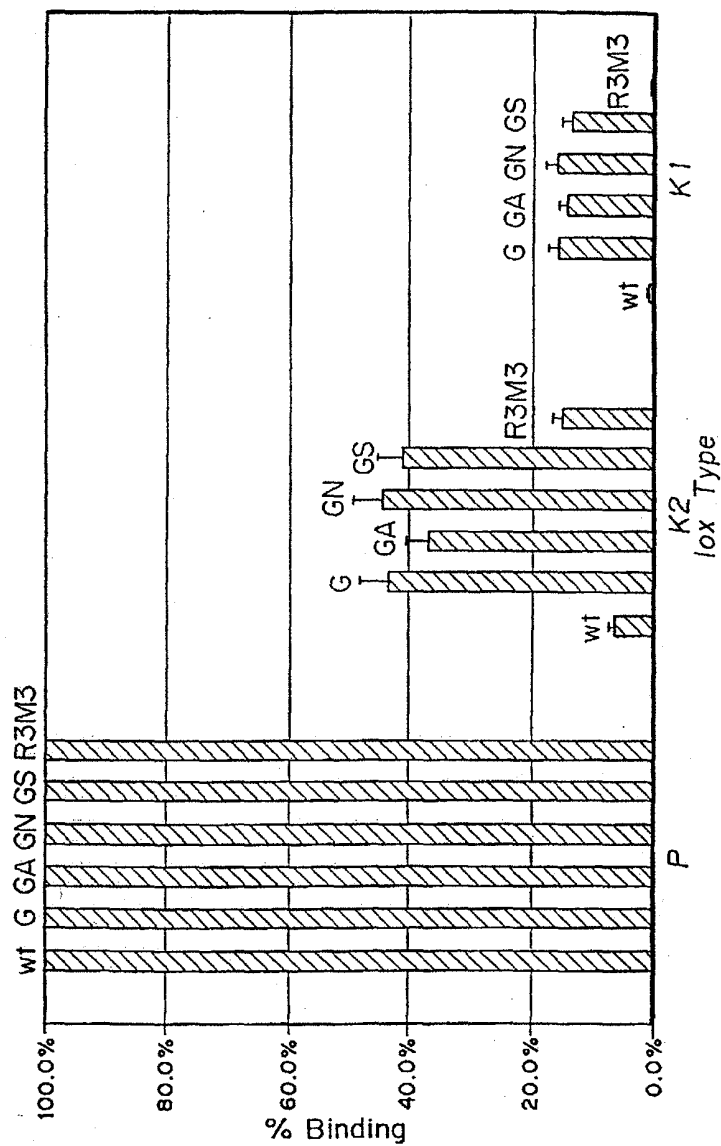
R3M3 Cre	97	90	95	30	82
E262G/D29A Cre	95	47	80	2	30
E262G/T316S Cre	98	32	81	1	12
E262G/D189N Cre	93	22	71	1	17
E262G Cre	90	28	78	2	20
wt Cre	95	0	18	0	5
	loxP <sup>2</sup>	loxK2 <sup>2</sup>	loxP- loxK2	loxK1 <sup>2</sup>	loxP- loxK1

FIG. 15

% Cre-mediated  $\text{lox}^2$  recombination

	97	99	92	82	56	67	90
R3M3 Cre	97	99	92	82	56	67	90
E262G/D29A Cre	95	95	82	37	35	34	47
E262G/T316S Cre	98	97	89	21	31	14	32
E262G/D189N Cre	93	98	68	6	9	5	22
E262G Cre	90	97	65	20	32	10	28
wt Cre	95	92	25	0	0	0	0
<div> <div>TT</div> <div>TG</div> <div>GT</div> <div>GG</div> <div>TC</div> <div>CC</div> <div>AA</div> </div>							
<div> <div><math>\text{loxP}</math> Halfsite</div> <div>ATAACTTCGTATA</div> </div>							

FIG. 16



wt FRT: GAAGTTCCTATTC TCTAGAA GTATAGGAAC TTC

Note - alternatively, a full site can also be used  
Having 3 repeat elements:

GAAGTTCCTATTC CGAAGTTCCTATTC TCTAGAA GTATAGGAAC TTC

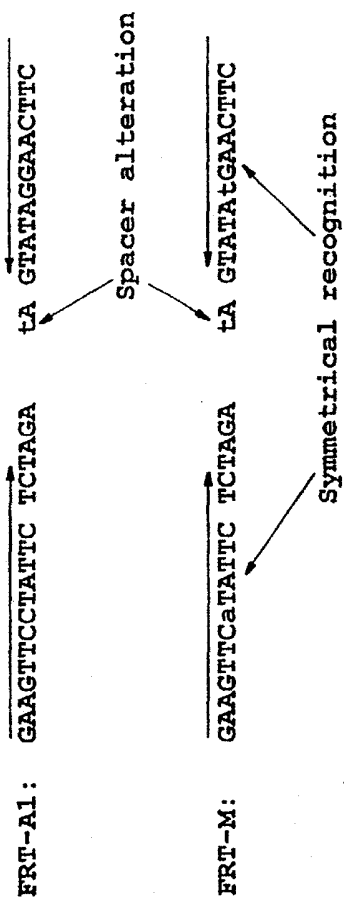
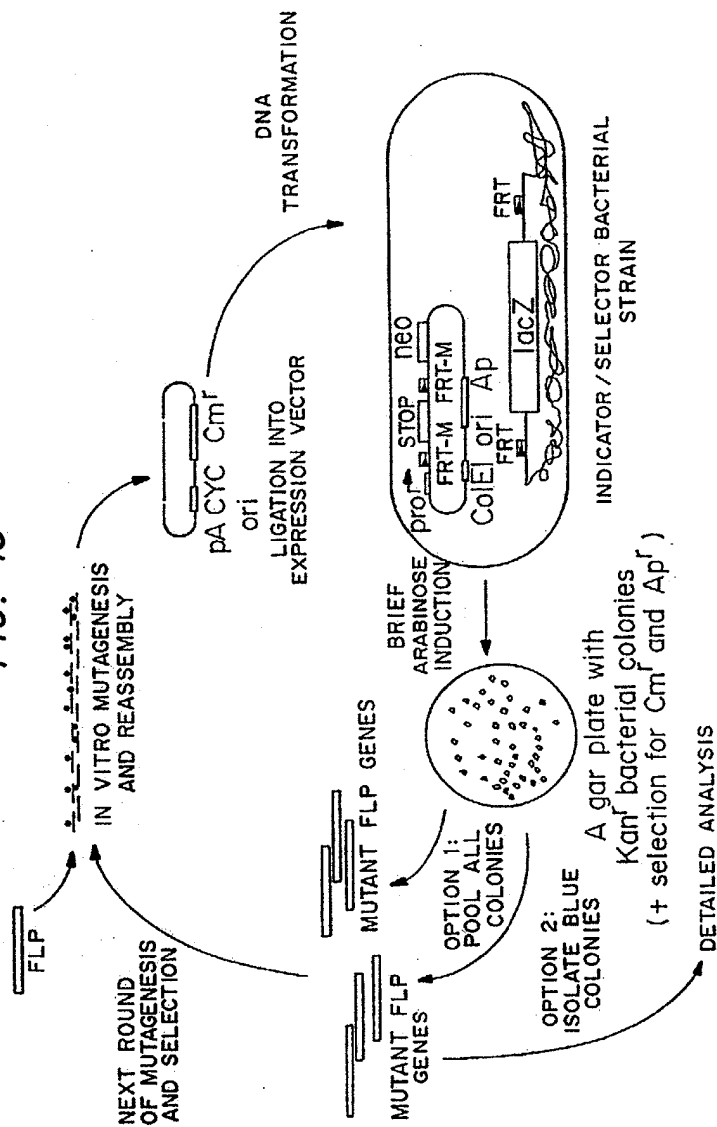


FIG. 17

FIG. 18



wt FRT: GAAGTTCCTATTC TCTAGAAA GTATAGGAACTTC

Note - alternatively, a full site can also be used  
Having 3 repeat elements:

GAAGTTCCTATTC CCGAAGTTCCTATTC TCTAGAAA GTATAGGAACTTC

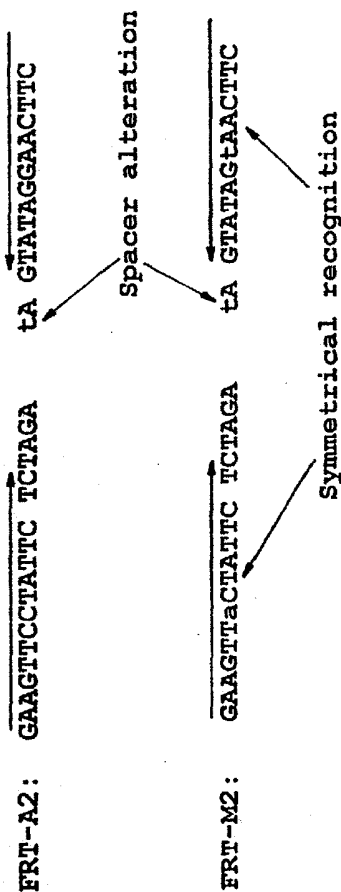


FIG. 19